Left column

RESULT 4

```
LOCUS       AF193855                2680 bp    mRNA    linear   PRI 05-DEC-2000
DEFINITION  Homo sapiens zinc finger protein of cerebellum ZIC2 (ZIC2) mRNA,
            complete cds.
ACCESSION   AF193855
VERSION     AF193855.1  GI:11065969
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2680)
AUTHORS     Yang,Y., Hwang,C.K., Junn,E., Lee,G. and Mouradian,M.M.
TITLE       ZIC2 and Sp3 Repress Sp1-induced Activation of the Human D1A
            Dopamine Receptor Gene
JOURNAL     J. Biol. Chem. 275 (49), 38863-38869 (2000)
PUBMED      10984499
REFERENCE   2  (bases 1 to 2680)
```

/translation="MLLDAGPQFPAIGVGSFARHHHHSAAAAAAAAEMQDRELSLAA
AQNGFVDSAAAHMGAFKLNPGAHELSPGQSSAFTSQGPGAYPGSAAAAAAALGPHA
AHVGSYSGPPFNSTRDFLFRSARLPGTSAPGGQHGLFSPGAGGLHHAHSDAQGHLLF
PGLPEOHGPHGSQNVLNGOMRLGLPGEVFGRSEQYRQVASPRTDPYSAAQLHNOYGPM
NMNMGMNMAAAAAHHHHHHHHPGAFFRYMRQQICIKQELICKWIDPEQLSNPKKSCNK
TFSTMHELVTHVSVEHVGVGPEQSNNVCFWEECPREGKPFKAKYKLVNHIRVHTGEKPF
PCPPPGCGKVFARSENLKIHKRTHTGEKPFQCEFEGCDRRFANSSDRKHMHVHTSDK
PYLCKMCDKSYTHPSSLRKHMKVHESSPQGSESSPAASSGYESSTPPGLVSPSAEPOS
SSNLSPAAAAAAAAAAAAAVSAVHRGGSGSGSGSGSGGAGGSGSGGGSGGGGGAGGGG
GSSGGSGTAGGHSGLSSNFNEWYV"

```
BASE COUNT      275 a    578 c    536 g    213 t
ORIGIN
```

Query Match

```
Best Local Similarity    100.0%;  Score 500; DB 9;  Length 1602;
Matches  500;  Conservative  0;  Pred. No. 7.2e-51;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy     1 ATGCTCCTGGACGCGGGTCCGCAGTTCCCGGCCATCGGGGTGGGCAGCTTCGCGCGCCAC   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGCTCCTGGACGCGGGTCCGCAGTTCCCGGCCATCGGGGTGGGCAGCTTCGCGCGCCAC   60

Qy    61 CATCACCACTCCGCGCCGCCGGGGGGGCCGCCGCTGCCGCAGATGCAGGACCGTGAACTG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CATCACCACTCCGCGCCGCCGGGGGGGCCGCCGCTGCCGCAGATGCAGGACCGTGAACTG  120

Qy   121 AGCCTGGCGGCGGCGCAGAACGGCTTCGTTGATTCCGCCGCGCCACATGGGAGCCTTC    180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGCGGCGGCGCAGAACGGCTTCGTTGATTCCGCCGCGCCACATGGGAGCCTTC    180

Qy   181 AAGCTCAACCCGGGCCAGCAGAGCTGTCCCCGGGCCAGCTCGCGCGTTCACGTCGCAG    240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGCTCAACCCGGGCCAGCACCAGAGCTGTCCCCGGGCCAGCTCGCGCGTTCACGTCGCAG  240

Qy   241 GGCCCCGGCGCGCTACCCCGGGCTCCCGGTGCGGCCAGCTTCCGCGCGCCAC         300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGCCCCGGCGCGCTACCCCGGGGTTCCGGTGCGGCCAGCTTCGCGCGCCAC         300

Qy   301 CACGCGCCGCGCACGTTGGCTCCTACTCTCGGGCGGGCGGCGGCGGGGGACTTCCTG    360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CACGCGCCGCGCACGTTGGCTCCTACTCTCGGGCGGGCGGCGGCGGGGGACTTCCTG    360

Qy   361 TTCCGCACGGCGCGGCGCGGGGGACTTCGGGCGCGGGGCGGGGGACTTCCTG        420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTCCGCAGCGCGCGGCGGGGGACTTCGGGCGCGGGGAGCACAGGGCGTGTTC        420

Qy   421 GGCGCCGCGGCGCTGCACCACCGCGCACTCGGACGCGCCGAGGGCCACCTCCTCTTC    480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGGCGCCGCGGCGCCTGCACCACCTCGGACGCGCCGAGGGCCACCTCCTCTTC        480

Qy   481 CCGGGGCCTGCCAGAGCAGCA         500
         ||||||||||||||||||||||
Db   481 CCGGGGCCTGCCAGAGCAGCA         500
```

Right column

```
AUTHORS     Yang,Y. and Mouradian,M.M.
TITLE       Direct Submission
JOURNAL     Submitted (11-OCT-1999) NINDS, NIH, 10 Center Drive, Bethesda, MD
            20892, USA
FEATURES             Location/Qualifiers
     source          1..2680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1..2680
                     /gene="ZIC2"
     CDS             1..1599
                     /gene="ZIC2"
                     /note="transcription factor; involved in cerebellar
                     development"
                     /codon_start=1
                     /product="zinc finger protein of cerebellum ZIC2"
                     /protein_id="AAG28409.1"
                     /db_xref="GI:11065970"
                     /translation="MLLDAGPQFPAIGVGSFARHHHHSAAAAAAAAEMQDRELSLAA
                     AQNGFVDSAAAHMGAFKLNPGAHELSPGQSSAFTSQGPGAYPGSAAAAAAALGPHA
                     AHVGSYSGPPFNSTRDFLFRSRFGPGDSAPCGGQHGLFSPGAGGLHHAHSDAQGHLLFP
                     GLPEOHGPHGSQNVLNGOMRLGLPGEVFGRSEQYRQVASPRTDPYSAAQLHNOYGPMN
                     MNMGMNMAAAAAHHHHHHHHPGAFFRYMRQQICIKQELICKWIDPEQLSNPKKSCNKT
                     FSTMHELVTHVSVEHVGVGPEQSNNVCFWEECPREGKPFKAKYKLVNHIRVHTGEKPF
                     CPPPGCGKVFARSENLKIHKRTHTGEKPFQCEFEGCDRRFANSSDRKHMHVHTSDKP
                     YLCKMCDKSYTHPSSLRKHMKVHESSPQGSESSPAASSGYESSTPPGLVSPSAEPOSS
                     SNLSPAAAAAAAAAAAVSAVHRGGSGSGSGSGSGGSGSGGGSGGGGGGAGGGGG
                     SSGGSGTAGGHSGLSSNFNEWYV"

BASE COUNT      572 a    798 c    723 g    586 t      1 others
ORIGIN
```

Query Match

```
Best Local Similarity    93.5%;  Score 467.6; DB 9;  Length 2680;
Matches  493;  Conservative  0;  Pred. No. 4.2e-47;  Mismatches  4;  Indels  3;  Gaps  2;
```

```
Qy     1 ATGCTCCTGGACGCGGGTCCGCAGTTCCCGGCCATCGGGGTGGGCAGCTTCGCGCGCCAC   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGCTCCTGGACGCGGGTCCGCAGTTCCCGGCCATCGGGGTGGGCAGCTTCGCGCGCCAC   60

Qy    61 CATCACCACTCCGCGCCGCCGGGGGGGCCGCCGCTGCCGCAGATGCAGGACCGTGAACTG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CATCACCACTCCGCGCCGCCGGGGGGGCCGCCGCTGCCGCAGATGCAGGACCGTGAACTG  120

Qy   121 AGCCTGGCGGCGGCGCAGAACGGCTTCGTTGATTCCGCCGCGCCACATGGGAGCCTTC    180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCTGGCGGCGGCGCAGAACGGCTTCGTTGATTCCGCCGCGCCACATGGGAGCCTTC    180

Qy   181 AAGCTCAACCCGGGCCAGCAGAGCTGTCCCCGGGCCAGCTCGCGCGTTCACGTCGCAG    240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGCTCAACCCGGGCCAGCAGAGCTGTCCCCGGGCCAGCTCGCGCGTTCACGTCGCAG    240

Qy   241 GGCCCCGGCGCGCTACCCCGGGCTCCCGGTGCGGCCAGCTTCGCGCGCCAC         300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGCCCCGGCGCGCTACCCCGGGCTCCCGGTGCGGCCAGCTTCGCGCGCCAC         300

Qy   301 CACGCGCCGCGCACGTTGGCTCCTACTCTCGGGCGGGCGGCGGCACGGGGGACTTCCTG  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CACGCGCCGCGCACGTTGGCTCCTACTCTCGGGCGGGCGGCGGCACGGGGGACTTCCTG  360

Qy   361 TTCCGCAGCGCGGCGCGGGGGACTTCGGGCGCGGGGCGGGGGACTTCCTG        420
         ||||||||| ||||||||| ||||||||||||||||||||||||||||||
Db   361 TTCCGCAGC-CGCGGCTTC--GGGGACTCGGCGCGCGGGCGGCGCGTGTTC        417

Qy   421 GGCGCCGCGGCGCTGCACCACCGCGCACTCGGACGCGCCGAGGGCCACCTCCTCTTC    480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   418 GGGCGCCGGCGGCCCTCCACCACCGCGCACCGCGCCGCAGGGCCACCTCCTCTTC    480

Qy   481 CCGGGGCCTGCCAGAGCAGCA         500
         |||||||||||||||||||||
Db   478 CCGGGGCCTGCCAGAGCAGCA         497
```

```
Qy    421 CCTGCCCAGACACCACTATGCCCCGTGGCCACCCGCCCTCCAGCTTGCCACCAAAGACCT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTGCCCAGACACCACTATGCCCCGTGGCCACCCGCCCTCCAGCTTGCCACCAAAGACCT 480

Qy    481 GCGACTTTGCTCAGGACTCC 500
          ||||||||||||||||||||
Db    481 GCGACTTTGCTCAGGACTCC 500
```

RESULT 3
AC011553/c

LOCUS   AC011553   35848 bp   DNA   linear   HTG 23-APR-2001
DEFINITION   Homo sapiens chromosome 19 clone LLNLR-277D11, WORKING DRAFT
              SEQUENCE, 3 ordered pieces.
ACCESSION   AC011553
VERSION   AC011553.3 GI:7711540
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 35848)
AUTHORS   DOE Joint Genome Institute.
TITLE   Sequencing of Human Chromosome 19
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 35848)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On May 6, 2000 this sequence version replaced gi:7690187.
              -----Genome Center
              Center: Joint Genome Institute
              Center code: JGI
              Web site: http://www.jgi.doe.gov

              Project Information
              Center Project Name: 26813, R31543
              Center clone name: LLNLR_277D11

              Summary Statistics
              Consensus quality: 35124 bases at least Q40
              Consensus quality: 35398 bases at least Q30
              Consensus quality: 35534 bases at least Q20
              Estimated insert size: 32650; agarose-fp estimation
              Estimated insert size: 35748; sum-of-contigs estimation
              Quality coverage: 10.55 in Q20 bases; agarose-fp estimation
              Quality coverage: 9.64 in Q20 bases; sum-of-contigs estimation.
              NOTE: This is a 'working draft' sequence. It currently
              * consists of 3 contigs. Gaps between the contigs
              * are represented as runs of N. The order of the pieces
              * is believed to be correct as given, however the sizes
              * of the gaps between them are based on estimates that have
              * provided by the submitter.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
              *   1   18613: contig of 18613 bp in length
              * 18614   18713: gap of unknown length
              * 18714   19277: contig of 564 bp in length
              * 19278   19377: gap of unknown length
              * 19378   35848: contig of 16471 bp in length.
FEATURES   Location/Qualifiers
     source   1..35848
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone_lib="Lawrence Livermore human cosmid library LLNLR"
                     /clone="LLNLR-277D11"
BASE COUNT   7758 a 11001 c 10770 g   6119 t   200 others
ORIGIN
```

```
Query Match                  44.4%;  Score 221.8;  DB 2;  Length 35848;
Best Local Similarity        91.4%;  Pred. No. 2.3e-25;
Matches  235;  Conservative  0;  Mismatches  22;  Indels  0;  Gaps  0;

Qy     73 CCGGACCCGGGCCGCCGAGATCATGATGCTGCCCGCCACCCGCCACCACGGAGAAGC 132
          ||  |  ||||||| |||| |||||||||||||| ||||||||||||||||||||||
Db   5256 CCCGTGCTCCCTGTCCAGATCATGATGGCTCGCGCCACCCGCCACCACGGAGAAGC 5197

Qy    133 CCAGATAGACGCCCCCCGGGGCGGCCCCCGGGGTCCCGGCGCCGCCCGGCCGAG 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5196 CCAGATAGACGCCCCCCGGGGCGGCCCCCGGGGTCCTGGAGTCCCGGCGCCGCCCGGCCGAG 5137

Qy    193 GACCCCCACCCCGCCGCCTGCGCCGCCGATGCTTGCAGTGCCCGCCCCATGGACAGGGATTACC 252
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5136 GACCCCCACCCCGCCGCCTGCGCCGCCGATGCTTGCAGTGCCCGCCCCATGGACAGGGATTACC 5077

Qy    253 CGCAGCATGAAACCCCCCGCCGGGGGCCAGCCTCCTGTACAGCCCCGCCCCTGCAGAGCG 312
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5076 CGCAGCATGAAACCCCCCGCCGGGGGCCAGCCTCCTGTACAGCCCCGCCCCTGCAGAGCG 5017

Qy    313 CCATGCTGCACTGCCCC 329
          | |  |||| |||||
Db   5016 GCCCCACCCCGGCCCC 5000
```

RESULT 4
AC093067/c

LOCUS   AC093067   6153 bp   DNA   linear   PRI 24-OCT-2001
DEFINITION   Homo sapiens chromosome 19 clone LLNLR-240D7, complete sequence.
ACCESSION   AC093067
VERSION   AC093067.2 GI:16356878
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 6153)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (09-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
REFERENCE   2 (bases 1 to 6153)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (24-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT   On Oct 24, 2001 this sequence version replaced gi:15144408.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www-shgc.stanford.edu
              Quality: Phrap Quality >~40 98.4% of Sequence;
              Estimated Total Number of Errors is 0.
              NOTE: This sequence is not the entire sequence of the clone. It
              is sequence generated to span the gap between AC005390 and
              AC011553. The overlap with AC005390 is 1172bp and the overlap with
              AC011553 is 1062bp. The sequence was finished by the Stanford
              Human Genome Center and Los Alamos National Laboratory.
FEATURES   Location/Qualifiers
     source   1..6153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="LLNLR-240D7"
BASE COUNT   1346 a   2032 c   1887 g   888 t
ORIGIN

Query Match   22.9%; Score 114.4; DB 9; Length 6153;
Best Local Similarity   99.1%; Pred. No. 8.7e-09;
```

RESULT 10

```
Qy  121 CACCTTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCCCCACCACTGCCCCTTTCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db  121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCCCCACCACTGCCCCTTTCC 180

Qy  181 TTTTTCTGACCTCCTTTTTGGAGGGGTCAGCGCCGCTGCCCAGACCATAGAGAGAGAGG 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTTTTCTGACCTCCTTTTTGGAGGGGTCAGCGCCGCTGCCCAGACCATAGAGAGAGAGG 240

Qy  241 CTCAGTTCCTGGGCGTTGCTGTTCTGCAGCCGCGCGTTTGGGGTGGCTCCAGTGAAGCCTCTC 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCAGTTCCTGGGCGTTGCTGTTCTGCAGCCGCGCGTTTGGGGTGGCTCCAGTGAAGCCTCTC 300

Qy  301 AGGCCAGGGGGCTGAGGTCCCGGTGGTGGTGGGGCCCAGGAGGAGGGGTCACTTGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGGCCAGGGGGCTGAGGTCCCGGTGGTGGTGGGGCCCAGGAGGAGGGGTCACTTGGC 360

Qy  361 GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGCAGGGGTCACTTGGC 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGCAGGGGTCACTTGGC 420

Qy  421 AGCATCAGCCAGACAGTGGCCCCGCCCGCGCCCGGGCCCATCCCCTGGCCCCCCGGCC 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGCATCAGCCAGACAGTGGCCCCGCCCGCGCCCGGGCCCATCCCCTGGCCCCCCGGCC 480

Qy  481 CTCACCCGGCGGGGCGCCCTCC 500
        ||||||||||||||||||||||
Db  481 CTCACCCGGCGGGGCGCCCTCC 500
```

LOCUS   A81356   2279 bp   DNA   linear   PAT 21-JAN-2000
DEFINITION Sequence 1 from Patent EP0900841.
ACCESSION   A81356
VERSION   A81356.1 GI:6731677
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2279)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE   LAG-3 splice variants
  JOURNAL   Patent: EP 0900841-A 1 10-MAR-1999;
          APPLIED RESEARCH SYSTEMS (AN); INST NAT SANTE RECH MED (FR)
FEATURES   Location/Qualifiers
     source   1..2279
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
BASE COUNT   360 a   832 c   589 g   498 t
ORIGIN

Query Match   100.0%; Score 500; DB 6; Length 2279;
Best Local Similarity 100.0%; Pred. No. 2.2e-95;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 TCAGGCTGCCTGAATCTGCCCAGCTTTCCAGCTTTCCTTCGTCGGCGCTCTGGTCATC 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TCAGGCTGCCTGAATCTGCCCAGCTTTCCAGCTTTCCTTCGTCGGCGCTCTGGTCATC 60

Qy   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCTCCCCTGTGGTCTCCCCTTCTTCTAGAACCCCTTCCCTC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCTCCCCTGTGGTCTCCCCTTCTTCTAGAACCCCTTCCCTC 120

Qy  121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCCCCACCACTGCCCCTTTCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCCCCACCACTGCCCCTTTCC 180

Qy  181 TTTTTCTGACCTCCTTTTTGGAGGGGTCAGCGCCGCTGCCCAGACCATAGAGAGAGAGG 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTTTTCTGACCTCCTTTTTGGAGGGGTCAGCGCCGCTGCCCAGACCATAGAGAGAGAGG 240

Qy  241 CTCAGTTCCTGGGCGTTGCTGTTCTGCAGCCGCGCGTTTGGGGTGGCTCCAGTGAAGCCTCTC 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCAGTTCCTGGGCGTTGCTGTTCTGCAGCCGCGCGTTTGGGGTGGCTCCAGTGAAGCCTCTC 300

Qy  301 AGGCCAGGGGGCTGAGGTCCCGGTGGTGGTGGGGCCCAGGAGGAGGGGTCACTTGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGGCCAGGGGGCTGAGGTCCCGGTGGTGGTGGGGCCCAGGAGGAGGGGTCACTTGGC 360

Qy  361 GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGCAGGGGTCACTTGGC 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGCAGGGGTCACTTGGC 420

Qy  421 AGCATCAGCCAGACAGTGGCCCCGCCCGCGCCCGGGCCCATCCCCTGGCCCCCCGGCC 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGCATCAGCCAGACAGTGGCCCCGCCCGCGCCCGGGCCCATCCCCTGGCCCCCCGGCC 480

Qy  481 CTCACCCGGCGGGGCGCCCTCC 500
        ||||||||||||||||||||||
Db  481 CTCACCCGGCGGGGCGCCCTCC 500
```

RESULT 11

LOCUS   AX001576   2279 bp   DNA   linear   PAT 10-MAR-2000
DEFINITION Sequence 1 from Patent WO9858059.
ACCESSION   AX001576
VERSION   AX001576.1 GI:7241708
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2279)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE   LAG-3 SPLICE VARIANTS
  JOURNAL   Patent: WO 9858059-A 1 23-DEC-1998;
          INST NAT SANTE RECH MED (FR); ROUSSY INST GUSTAVE (FR)
FEATURES   Location/Qualifiers
     source   1..2279
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
BASE COUNT   360 a   832 c   589 g   498 t
ORIGIN

Query Match   100.0%; Score 500; DB 6; Length 2279;
Best Local Similarity 100.0%; Pred. No. 2.2e-95;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 TCAGGCTGCCTGAATCTGCCCAGCTTTCCAGCTTTCCTTCGTCGGCGCTCTGGTCATC 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TCAGGCTGCCTGAATCTGCCCAGCTTTCCAGCTTTCCTTCGTCGGCGCTCTGGTCATC 60

Qy   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCTCCCCTGTGGTCTCCCCTTCTTCTAGAACCCCTTCCCTC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCTCCCCTGTGGTCTCCCCTTCTTCTAGAACCCCTTCCCTC 120

Qy  121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCCCCACCACTGCCCCTTTCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCCCCACCACTGCCCCTTTCC 180

Qy  181 TTTTTCTGACCTCCTTTTTGGAGGGGTCAGCGCCGCTGCCCAGACCATAGAGAGAGAGG 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTTTTCTGACCTCCTTTTTGGAGGGGTCAGCGCCGCTGCCCAGACCATAGAGAGAGAGG 240

Qy  241 CTCAGTTCCTGGGCGTTGCTGTTCTGCAGCCGCGCGTTTGGGGTGGCTCCAGTGAAGCCTCTC 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCAGTTCCTGGGCGTTGCTGTTCTGCAGCCGCGCGTTTGGGGTGGCTCCAGTGAAGCCTCTC 300

Qy  301 AGGCCAGGGGGCTGAGGTCCCGGTGGTGGTGGGGCCCAGGAGGAGGGGTCACTTGGC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGGCCAGGGGGCTGAGGTCCCGGTGGTGGTGGGGCCCAGGAGGAGGGGTCACTTGGC 360

Qy  361 GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGCAGGGGTCACTTGGC 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
BASE COUNT      166 a      261 c      246 g       72 t        3 others
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 748;
Best Local Similarity   100.0%;  Pred. No. 1.8e-36;
Matches   500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGTCTACGGCATTGCTGAGGACGCGTCGCCCAGGCATCGCTAATGAGGACGCGCCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGCATTGCTGAGGACGCGTCGCCCAGGCATCGCTAATGAGGACGCGCCAGG 60

Qy     61 GCATCGCTAATGAGGACACCACCACCAGTGCATCGCCAACGAGGACGCCCAGGGCATCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATCGCTAATGAGGACACCACCACCAGTGCATCGCCAACGAGGACGCCCAGGGCATCG 120

Qy    121 CCGAGGACGCCCATCCAGGGCATCGGCCCAGGGCATCGCCCAATGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGGACGCCCATCCAGGGCATCGGCCCAGGGCATCGCCCAATGGGG 180

Qy    181 TCGCGCGCACAGGGCATCGCCAATGAGGACGCCACCACCAGGGCATCGCCAACTGGGACGCCG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGCGCGCACAGGGCATCGCCAATGAGGACGCCACCACCAGGGCATCGCCAACTGGGACGCCG 240

Qy    241 TCCACGGCTTCGCCAACGGGGACGCCGTCTCAGCTTCGCCAACGGGGACGCGGCCCAGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCACGGCTTCGCCAACGGGGACGCCGTCTCAGCTTCGCCAACGGGGACGCGGCCCAGG 300

Qy    301 GCATCGCCAACGGGGACGCCACCAAGGGCATGGCAGAGGTCACCATCCACGGCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCATCGCCAACGGGGACGCCACCAAGGGCATGGCAGAGGTCACCATCCACGGCATCG 360

Qy    361 CTAACGAGGACGCCGTCCAGGGCATCCTACGAGGTGGCCCGCCAGGGCATGCCAACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTAACGAGGACGCCGTCCAGGGCATCCTACGAGGTGGCCCGCCAGGGCATGCCAACG 420

Qy    421 AGGACGCCGCCCAGGGAAATCGCCGAGGATGTGCCACAGGGACGCCGCCAACG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGACGCCGCCCAGGGAAATCGCCGAGGATGTGCCACAGGGACGCCGCCAACG 480

Qy    481 CCCAGGGCATCGCCAACAAG 500
          ||||||||||||||||||||
Db    481 CCCAGGGCATCGCCAACAAG 500
```

RESULT 2
LOCUS   AX201594   1201 bp   DNA   linear   PAT 30-AUG-2001
DEFINITION   Sequence 15 from Patent WO0153349.
ACCESSION   AX201594
VERSION   AX201594.1 GI:15391443
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
             Chen,Y.T.
  TITLE   Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 15 26-JUL-2001;
             LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
             SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
             INC. (US)
FEATURES   Location/Qualifiers
     source   1..1201
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

BASE COUNT   255 a   377 c   349 g   178 t   42 others
ORIGIN

Query Match   100.0%; Score 500; DB 6; Length 1201;
Best Local Similarity   100.0%; Pred. No. 1.6e-36;
Matches   500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 3
LOCUS   HSM801408   2448 bp   mRNA   linear   PRI 18-FEB-2000
DEFINITION   Homo sapiens mRNA; cDNA DKFZp434C196 (from clone DKFZp434C196);
             partial cds.
ACCESSION   AL133561
VERSION   AL133561.1 GI:6599133
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2448)
  AUTHORS   Koehrer,K., Beyer,A.., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-DEC-1999) MIPS, Am Klopferspitz 18a, D-82152
             Martinsried, GERMANY
  COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
             sequenced by BMFZ (Biomedical Research Center at the Charite,
             Berlin/Germany) within the cDNA sequencing consortium of the German
             Genome Project.
             This clone (DKFZp434C196) is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
             information about the clone and the sequencing project is available
             at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES   Location/Qualifiers
     source   1..2448
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434C196"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"

```
Qy      1 GAGTCTACGGCATTGCTGAGGACGCGTCGCCCAGGCATCGCTAATGAGGACGCGCCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGCATTGCTGAGGACGCGTCGCCCAGGCATCGCTAATGAGGACGCGCCAGG 60

Qy     61 GCATCCCTAATGAGGACACCACCACCAGTGCATCGCCAACGAGGACGCCCAGGGCATCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCATCGCTAATGAGGACACCACCACCAGTGCATCGCCAACGAGGACGCCCAGGGCATCG 60

Qy    121 CCGAGGACGCCCATCCAGGGCATCGCCAACGAGGACGCCCAGGGCATCGCCCAATGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCGAGGACGCCCATCCAGGGCATCGCCAACGAGGACGCCCAGGGCATCGCCCAATGGGG 120

Qy    181 TCGCGCGCACAGGGCATCGCCAATGAGGACGCCACCACCAGGGCATCGCCAACTGGGACGCCG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCGCGCGCACAGGGCATCGCCAATGAGGACGCCACCACCAGGGCATCGCCAACTGGGACGCCG 180

Qy    241 TCCACGGCTTCGCCAACGGGGACGCCGTCCTCCAGCTTCGCCAACGGGGACGCGGCCCAGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCCACGGCTTCGCCAACGGGGACGCCGTCCTCCAGCTTCGCCAACGGGGACGCGGCCCAGG 240

Qy    301 GCATCGCCAACGGGGACGCCACCAAGGGCATGGGCAGAGGTCACCATCCACGGCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCATCGCCAACGGGGACGCCACCAAGGGCATGGGCAGAGGTCACCATCCACGGCATCG 300

Qy    361 CTAACGAGGACGCCGTCCAGGGCATCCTACGAGGTGGCCCGCCAGGGCATGCCAACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTAACGAGGACGCCGTCCAGGGCATCCTACGAGGTGGCCCGCCAGGGCATGCCAACG 360

Qy    421 AGGACGCCGCCCAGGGAATCGCCGAGGATGTCGGCACAGGGACGCCGCCAACG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGGACGCCGCCCAGGGAATCGCCGAGGATGTCGGCACAGGGACGCCGCCAACG 420

Qy    481 CCCAGGGCATCGCCAACAAG 500
          ||||||||||||||||||||
Db    481 CCCAGGGCATCGCCAACAAG 500
```

RESULT 4

```
Query Match 99.7%; Score 498.4; DB 9; Length 2448;
Best Local Similarity 99.8%; Pred. No. 1.7e-36;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAGTCTACGCATTGCTGAGGACGCGCTCCCCAGGGCATCGCTAATGAGGACGCCGACCAGG 60
          ||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    167 GAGTCTACGGCATTGCTGAGGACGGCTCCCCAGGGCATCGCTAATGAGGACGCCGACCAGG 226

Qy     61 GCATCGCTAATGAGGACACCACCCAGGGCATCCCCAACGAGGAAGCCGCCCAGGGCATCG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    227 GCATCGCTAATGAGGACACCACCCAGGGCATCCCCAACGAGGAAGCCGCCCAGGGCATCG 286

Qy    121 CCGAGGACGACGCCATCCAGGGCATCGCCAACGAGGAGGTTGCCCAGGGCATCGCCAATGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 CCGAGGACGACGCCATCCAGGGCATCGCCAACGAGGAGGTTGCCCAGGGCATCGCCAATGGGG 346

Qy    181 TCGCCCGCACAGGGCATCGCCAATGAGGACGACGCCACCCAGGGCATCGCCAACGACGTCG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 TCGCCCGCACAGGGCATCGCCAATGAGGACGACGCCACCCAGGGCATCGCCAACGACGTCG 406

Qy    241 TCCACCGGCTTCGCCCACACGGGGACGCGCCCTCAGCTTCGGCCACACGGGGAGGACGCGCCCAGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 TCCACCGGCTTCGCCCACACGGGGACGCGCCCTCAGCTTCGGCCACACGGGGAGGACGCGCCCAGG 466

Qy    301 GCATCGCCAACGGGGAGGCGCCACCCAGGGCAATGGGCAACGAGGTCACCATCCACGGCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 GCATCGCCAACGGGGAGGCGCCACCCAGGGCAATGGGCAACGAGGTCACCATCCACGGCATCG 526

Qy    361 CTAACGAGGACGGCGTCCAGGGCATCGCTAACGAGGAGGTGGCCCAACGAGGTGGCCCAACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 CTAACGAGGACGGCGTCCAGGGCATCGCTAACGAGGAGGTGGCCCAACGAGGTGGCCCAACG 586

Qy    421 AGGACGCGCCCCAGGGAATCGCCGACGGATGTCGCACAGGGCCATCGCCACGAGGACGCCG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    587 AGGACGCGCCCCAGGGAATCGCCGACGGATGTCGCACAGGGCCATCGCCACGAGGACGCCG 646

Qy    481 CCCAGGGCCATCGCCAACAAG 500
          |||||||||||||||||||||
Db    647 CCCAGGGCCATCGCCAACAAG 666
```

```
     gene            1..1744
                     /gene="DKFZp434C196"
     CDS             <1..1744
                     /gene="DKFZp434C196"
                     /note="similar to mucins"
                     /codon_start=2
                     /product="hypothetical protein"
                     /protein_id="CAB63715.1"
                     /db_xref="GI:6599134"
                     /translation="SPSRASLTRTPSRASLMRRPSTASLRTRPSRASPTRMPSRASLK
                     MTFPRASITKMESTALLRTLPRASLMRRTPRASLMTGTPRASPTRKPPRASPRTPSRA
                     SPTRLPRASPMGSPHRASPMRTPPRASPTGTPSTASPTGTPSSASPTGTPPRASPTG
                     TPPRAWATRSPSTASLTRTPSRASLTRWPPRASPTRPPRESPRMSHRASPTRTPPRA
                     SPTRRPPRASPTTPPRESLRTSHRASPTRMPPRASPTGSPPRASPMTP
                     PRASPRTPPRASPTTTPPRESLRTSRASLMKMESTVSITRTPPRASP
                     TGTPSRASPTGTPSRASLTGSPSRASLIGTPSRASLIGTPSRASLTGTP
                     PRASLTGTSTASLTRTPSMASLTRTPQSSSLTRTPSMASLTRTPPRAS
                     LTRTPPRASLTRTPPRASLTRPSMVSLKRSPSRASLTRTPSRASLTRTPSRASLTRT
                     PSTASLTGTPSLTLTRTPPTASLTRTPPTASLTRMPSTASLTRKSNVN
                     QQCPASTPSSEVIS"
     polyA_signal    2397..2402
     polyA_site      2418

BASE COUNT      643 a    727 c    648 g    430 t
ORIGIN
```

```
RESULT 4
ACCESSION    AL356585
DEFINITION   Homo sapiens chromosome 13 clone RP11-341D18, *** SEQUENCING IN
             PROGRESS ***, 13 unordered pieces.
LOCUS        AL356585    184590 bp    DNA    linear    HTG  20-JAN-2001
AL356585/c

ORIGIN
BASE COUNT    643 a    727 c    648 g    430 t

Matches 499; Conservative  0; Mismatches  1; Indels  0; Gaps  0;
Best Local Similarity  99.8%;  Pred. No. 1.7e-36;
Query Match  99.7%;      Score 498.4; DB 9; Length 2448;

Qy      1 GAGTCTACGGCATTGCTGAGGAGCGCTGCCCAGGGCCATCGCTAATGAGGAGGACGCCAGG    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    167 GAGTCTACGGCATTGCTGAGGAGCGCTGCCCAGGGCCATCGCTAATGAGGAGGACGCCAGG   226

Qy     61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCCAGGGCCATCGCTAATGAGGAGGACGCCGACCAGG   120
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    227 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCCAGGGCCATCGCCAACGAGGAAGCCCCAGGGCATCG   286

Qy    121 CCGAGGAGGCCATCCAGGGCATCGCCCAACGAGGAGGAGGTTCCCCAGGGCATCGCCAATGGGG   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 CCGAGGACCCCATCCAGGGCATCGCCCAACGAGGAGGAGGTTGCCCAGGCATCGCCAATGGGG   346

Qy    181 TCGCCCGCACAGGGCATCGCCAATGAGGAGCGCGCCAACTGGGACGCCGCGCG   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 TCGCCCGCACAGGGCCATCGCCAATCGAGGACGCACCGCCAACTGGGACGCCG   406

Qy    241 TCCACGGCTTCGCCAACGGGGACGCGCGTCCTCACCTTCGGCCAACGGGGACGCCGCCGCG   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 TCCACGGCTTCGCCAACGGGGACGCCGCCGTCCTCAGCTTCGGCCAACGGGGACGCCCCAGG   466

Qy    301 GCATCGCCAACGGGGACGCCACCAAGGGCATGGCAACGAGGTCACCATCCACGGCATCG   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    467 GCATCGCCAACGGGGACGCCACCAAGGGCACGCAACGAGGTCACCATCCACGGCATCG   526

Qy    361 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCGCCGCCGCCCAACG   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCGCCGCCGCCCAACG   586

Qy    421 AGGACGCCGCCCCAGGGAATCGCCAGGAGTGTCGCACAGGACGCCACGAGGACGCCG   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    587 AGGACGCCGCCCCAGGGAATCGCCACAGGGCCATCGCCAACAGAGGACGCCG   646

Qy    481 CCCAGGGGCATCGCCAACAAG   500
          |||||||||||||||||||||
Db    647 CCCAGGGCATCGCCAACAG   666
```

```
polyA_signal   2397..2402
polyA_site     2418
               2397..2402
gene           1..1744
               /gene="DKFZp434C196"
CDS            <1..1744
               /gene="DKFZp434C196"
               /note="similar to mucins"
               /codon_start=2
               /product="hypothetical protein"
               /protein_id="CAB63715.1"
               /db_xref="GI:6599134"
               /translation="SPSRASLTRTPRASLMRRPSTASLTRTPSRASPTRMPSRASLK
               MTPRASLTKMESTALLRTLPRASLMRTPTRASLMRPPPRASPTRKPPRASPRTPSRA
               SPTRLPRASPMGSPHRASPMRTPPRASPTGTPSTASPTGPTPSSASPTGPTPRASPTG
               TPPRAWATRSPSTASLTRTPSRASLTRWPPRASPTRTPPRESPRMSHRASPTRTPPRA
               SPTRRPPRASPTRTPPRESLRHRASPTRTPSWASPTTPSRASLMKMESTVSITRTPPRASP
               TGTPSRASPTGTPSRASLTGTPSRASLIGTPSRASLIGTPSRASLTGTP
               PRASLTGTSSTASLTRRPSRASLTRTQSSSITRTPSMASLTRTPPRASLTRTPPRAS
               LRTPPRASLTRTPPRASLTRTPSMVSLKRSPSRASLTRTPSRASLIRTPPRASLTRTPPRAS
               PSTASLTGTPTPTASLTRTPIPTASLTRSPPTASLTRTPSTASLIRMPSTASLTRKSNVN
               QCPASTPSSEVIS"
               /dev_stage="adult"
```

BASE COUNT   253 a   348 c   348 g   136 t
ORIGIN
Best Local Similarity   100.0%;   Score 500; DB 6; Length 1085;
Matches 500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy     1 CACAGCGCCCGCATGTACACACATGATGGAGACGGAGCTGAAGCCGCCGCAACCGCCAGCAA   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CACAGCGCCCGCATGTACACACATGATGGAGACGGAGCTGAAGCCGCCGCAACCGCCAGCAA   60

Qy    61 ACTTCGGGGGGCCGCGGGGGGGCCCGGCCAACCTCCACCCGCCGGCCAACCAGAAAAAC  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACTTCGGGGGGCCGCGGGGGGGCCCGGCCAACCTCCACCCGCCGGCCAACCAGAAAAAC  120

Qy   121 AGCCCGGACCCGTCAAGCGGCCCCATGAATGCCTTCATGGGTGTGGGTCCCGGGGGCAGCGG  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCCGGACCCGTCAAGCGGCCCCATGAATGCCTTCATGGGTGTGGGTCCCGGGGGCAGCGG  180

Qy   181 CGCAAGAGATGGGCCCAGGAGAACCCCAAGATACACAACTCGGAGAGATCAGCAAGGCCCTGGGC  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CGCAAGAGATGGGCCCAGGAGAACCCCAAGATACACAACTCGGAGAGATCAGCAAGGCCCTGGGC  240

Qy   241 GCCGAGTGGAAAACTTTTTGTCCGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCG  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCCGAGTGGAAAACTTTTTGTCCGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCG  300

Qy   301 CTGCGAGCGCGCTGCACATGAAGGAGCACCCCGATTATAAAATACCGGCCCCCGGGGAAAACC  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTGCGAGCGCGCTGCACATGAAGGAGCACCCCGATTATAAAATACCGGCCCCCGGGGAAAACC  360

Qy   361 AAGACGGCTCATGAAGAAGGATAAGTACACGCCTGGCGGGCGGCTGCTGCCGCCGGGGCGGC  420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGACGGCTCATGAAGAAGGATAAGTACACGCCTGGCGGGCGGCTGCTGCCGCCGGGGCGGC  420

Qy   421 AATAGCATGGCGAGCGGGGTCGGGGGGTGGGCGGCGTGAACCACGCGC  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AATAGCATGGCGAGCGGGGTCGGGGGGTGGGCGGCGTGAACCACGCGC  480

Qy   481 ATGGACAGTTACGCGCACAT  500
         ||||||||||||||||||||
Db   481 ATGGACAGTTACGCGCACAT  500
```

RESULT 2
HSSOX2G
LOCUS   HSSOX2G   1085 bp   mRNA   linear   PRI 28-JUL-1995
DEFINITION H.sapiens sox-2 mRNA (partial).
ACCESSION   Z31560
VERSION   Z31560.1 GI:854181
KEYWORDS   sox-2 gene.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1085)
AUTHORS   Stevanovic,M., Zuffardi,O., Collignon,J., Lovell-Badge,R. and
            Goodfellow,P.
TITLE   The cDNA sequence and chromosomal location of the human SOX2 gene
JOURNAL   Mamm. Genome 5 (10), 640-642 (1994)
MEDLINE   95152171
REFERENCE   2 (bases 1 to 1085)
AUTHORS   Stevanovic.M.
TITLE   Direct Submission
JOURNAL   Submitted (23-MAR-1994) Stevanovic M., Institute of Molecular
            Genetics and Genetic Engineering, Vojvode Stepe 283, Belgrade,
            Serbia, Yugoslavia, 11000
FEATURES   Location/Qualifiers
     source   1..1085
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="FBCt1"

gene   /dev_stage="fetus"
                1..966
                /gene="sox-2"
CDS   <1..966
                /gene="sox-2"
                /citation=[1]
                /codon_start=1
                /protein_id="CAA83435.1"
                /db_xref="GI:854182"
                /db_xref="SWISS-PROT:P48431"
                /translation="HSARMYNNMETELKPPGPQQTSGGGGNSTAAAAGGNQKNSPDR
                VKRPMNAFMVWRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLR
                ALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQR
                MDSYAHMNGWSNGSYSMMQDDLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQ
                TYMNGSSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRD
                MISMYLPGAEVPEPAAPSRLHMSQHYQSGPVPGTAINGTLPLSHM"
```

BASE COUNT      253 a      348 c      348 g      136 t
ORIGIN
Best Local Similarity    100.0%;    Score 500;  DB 9;  Length 1085;
Matches 500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

```
Qy     1 CACAGCGCCCGCATGTACACACATGATGGAGACGGAGCTGAAGCCGCCGCAACCGCCAGCAA   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CACAGCGCCCGCATGTACACACATGATGGAGACGGAGCTGAAGCCGCCGCAACCGCCAGCAA   60

Qy    61 ACTTCGGGGGGCCGCGGGGGGGCCCGGCCAACCTCCACCCGCCGGCCAACCAGAAAAAC  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACTTCGGGGGGCCGCGGGGGGGCCCGGCCAACCTCCACCCGCCGGCCAACCAGAAAAAC  120

Qy   121 AGCCCGGACCCGTCAAGCGGCCCCATGAATGCCTTCATGGGTGTGGGTCCCGGGGGCAGCGG  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGCCCGGACCCGTCAAGCGGCCCCATGAATGCCTTCATGGGTGTGGGTCCCGGGGGCAGCGG  180

Qy   181 CGCAAGAGATGGGCCCAGGAGAACCCCAAGATACACAACTCGGAGAGATCAGCAAGGCCCTGGGC  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CGCAAGAGATGGGCCCAGGAGAACCCCAAGATACACAACTCGGAGAGATCAGCAAGGCCCTGGGC  240

Qy   241 GCCGAGTGGAAAACTTTTTGTCCGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCG  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCCGAGTGGAAAACTTTTTGTCCGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCG  300

Qy   301 CTGCGAGCGCGCTGCACATGAAGGAGCACCCCGATTATAAAATACCGGCCCCCGGGGAAAACC  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTGCGAGCGCGCTGCACATGAAGGAGCACCCCGATTATAAAATACCGGCCCCCGGGGAAAACC  360

Qy   361 AAGACGGCTCATGAAGAAGGATAAGTACACGCCTGGCGGGCGGCTGCTGCCGCCGGGGCGGC  420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGACGGCTCATGAAGAAGGATAAGTACACGCCTGGCGGGCGGCTGCTGCCGCCGGGGCGGC  420

Qy   421 AATAGCATGGCGAGCGGGGTCGGGGGGTGGGCGGCGTGAACCACGCGC  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AATAGCATGGCGAGCGGGGTCGGGGGGTGGGCGGCGTGAACCACGCGC  480

Qy   481 ATGGACAGTTACGCGCACAT  500
         ||||||||||||||||||||
Db   481 ATGGACAGTTACGCGCACAT  500
```

RESULT 3
HUMHMGBOX
LOCUS   HUMHMGBOX   1109 bp   mRNA   linear   PRI 08-MAY-1993
DEFINITION Homo sapiens (clone 6AR33) HMG box mRNA, 3' end cds.
ACCESSION   L07335
VERSION   L07335.1 GI:184239
KEYWORDS   HMG box.
SOURCE   Homo sapiens retina cDNA to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1109)
AUTHORS   Sadler,L.A., Badzioch,M.D., Wagner,M., Graves,K.A., Swaroop,A.,

BASE COUNT   924 a   1109 c   1110 g   943 t   5 others
ORIGIN

Query Match   100.0%;   Score 500;   DB 6;   Length 4091;
Best Local Similarity   100.0%;   Pred. No. 6.5e-51;
Matches 500; Conservative   0; Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 CCGGCCGTCTATGCTCCAGGCCCCCTCTCCTCTCGCGGTGAACCCGCCAGCCGCCGCCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CCGGCCGTCTATGCTCCAGGCCCCCTCTCCTCTCGCGGTGAACCCGCCAGCCGCCGCCG 60

Qy   61 ATGTACAGCATGATGATGGAGACCGACCTGCACTCGCCGGGTGCCGGCCCCCACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATGTACAGCATGATGATGGAGACCGACCTGCACTCGCCGGGTGCCGGCCCCCACG 120

Qy   121 AACCTCTCGGGCCCCCGCGGCGGCGGCGGCGGCGGCGGCCGGCGGCGGCGGCGGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AACCTCTCGGGCCCCCGCGGCGGCGGCGGCGGCGGCGGCCGGCGGCGGCGGCGGC 180

Qy   181 GGCGGGGGGCCCAAGGCCAACCAGGACACGGCGCCCATGAACGCCTTCATGGTG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCGGGGGGCCCAAGGCCAACCAGGACACGGCGCCCATGAACGCCTTCATGGTG 240

Qy   241 TGGTCCCGGGCCCGGCGGCCAGCGCGCGGCGCAAGATGGCCCCAGGAGAACCCCAGAGGAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGGTCCCGGGCCCGGCGGCCAGCGCGCGGCGCAAGATGGCCCCAGGAGAACCCCAGAGGAG 300

Qy   301 ATCAGCAAGCCCTGGGGCGGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATCAGCAAGCCCTGGGGCGGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTC 360

Qy   361 ATCGACGAGGCGGCCAACGGCTGCGCCGCTGCACTCGAAGGACACCCGGATTACAAGTAC 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ATCGACGAGGCGGCCAACGGCTGCGCCGCTGCACTCGAAGGACACCCGGATTACAAGTAC 420

Qy   421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGCCGGGCGGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGCCGGGCGGG 480

Qy   481 CTCCTGGCGCCGGGCGGCGGGGG 500
           |||||||||||||||||||||||
Db   481 CTCCTGGCGCCGGGCGGCGGGGG 500

RESULT 2
HSSOX1
LOCUS   HSSOX1   4091 bp   DNA   linear   PRI 09-FEB-2001
DEFINITION   Homo sapiens sox1 gene.
ACCESSION   Y13436
VERSION   Y13436.1 GI:4128158
KEYWORDS   SOX1 gene; Sry-related Box 1 protein.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4091)
AUTHORS   Malas,S., Duthie,S.M., Mohri,F., Lovell-Badge,R. and Episkopou,V.
TITLE   Cloning and mapping of the human SOX1: a highly conserved gene
             expressed in the developing brain
JOURNAL   Mamm. Genome 8 (11), 866-868 (1997)
MEDLINE   98051911
REFERENCE   2 (bases 1 to 4091)
AUTHORS   Malas,S.
TITLE   Direct Submission
JOURNAL   Submitted (28-MAY-1997) S. Malas, MRC Clinical Sciences Centre,
             Mouse Embryology, Du Cane Rd, London, W12 ONN, London, UK
REFERENCE   3 (bases 1 to 4091)
AUTHORS   Malas,S.
REMARK   Revised by [3]
JOURNAL   Direct Submission
             Submitted (28-MAY-1997) S. Malas, MRC Clinical Sciences Centre,
             Mouse Embryology, Du Cane Rd, London, W12 ONN, London, UK
COMMENT   On Jan 8, 1999 this sequence version replaced gi:2220882.

FEATURES   Location/Qualifiers
     source   1..4091
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="q33-34"
                     /clone="psxBg1.1"
     gene   61..1224
                     /gene="SOX1"
     CDS   61..1224
                     /gene="SOX1"
                     /codon_start=1
                     /product="Sry-related Box 1 protein"
                     /protein_id="CAA73847.1"
                     /db_xref="GI:2230883"
                     /db_xref="SWISS-PROT:O00570"
                     /translation="MYSMMMETDLHSPGGAQAPTNLSGPAGGGGGGGGGGGGPKANQD
                     RVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKRPFID
                     EAKRLRALHMKEHPDYKYRPRRKTKTLLKDKYSLAGGLLAAGAGGYPGVSVAAA
                     VGAAPVGQRLESPGGAAGAYAHVNGWANGAYPGSVAAAAAAMMQEAQLAYGQHPG
                     AGGAIPHRTEPAHPHPHPHPHNPQPMHRYDMGALQYSPISNSQGTMSASPSGYGGL
                     PYGAAAAAAAAHQNSAVAAAAAAAAAAASSGALGALSSLVKSEPSGSPPAPAHSRAPCPG
                     DLREMISMYLPAGEGDPAAAAAAAAQSRLHSLPQHYQGAGAGVNGTVPLHI"

Related sequences: AI279621, AI298071, AI215744, AA960996, D81624,
R24723, R20579, T07302, R14439, AA961095, T06325, R46080.

Related Sequences section / second alignment block:

BASE COUNT   924 a   1109 c   1110 g   943 t   5 others
ORIGIN

Query Match   100.0%;   Score 500;   DB 9;   Length 4091;
Best Local Similarity   100.0%;   Pred. No. 6.5e-51;
Matches 500; Conservative   0; Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 CCGGCCGTCTATGCTCCAGGCCCCCTCTCCTCTCGCGGTGAACCCGCCAGCCGCCGCCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CCGGCCGTCTATGCTCCAGGCCCCCTCTCCTCTCGCGGTGAACCCGCCAGCCGCCGCCG 60

Qy   61 ATGTACAGCATGATGATGGAGACCGACCTGCACTCGCCGGGTGCCGGCCCCCACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATGTACAGCATGATGATGGAGACCGACCTGCACTCGCCGGGTGCCGGCCCCCACG 120

Qy   121 AACCTCTCGGGCCCCCGCGGCGGCGGCGGCGGCGGCGGCCGGCGGCGGCGGCGGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AACCTCTCGGGCCCCCGCGGCGGCGGCGGCGGCGGCGGCCGGCGGCGGCGGCGGC 180

Qy   181 GGCGGGGGGCCCAAGGCCAACCAGGACACGGCGCCCATGAACGCCTTCATGGTG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCGGGGGGCCCAAGGCCAACCAGGACACGGCGCCCATGAACGCCTTCATGGTG 240

Qy   241 TGGTCCCGGGCCCGGCGGCCAGCGGCGCGGCGCAAGATGGCCCCAGGAGAACCCCAGAGGAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGGTCCCGGGCCCGGCGGCCAGCGGCGCGGCGCAAGATGGCCCCAGGAGAACCCCAGAGGAG 300

Qy   301 ATCAGCAAGCCCTGGGGCGGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATCAGCAAGCCCTGGGGCGGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTC 360

Qy   361 ATCGACGAGGCGGCCAACGGCTGCGCCGCTGCACTCGAAGGACACCCGGATTACAAGTAC 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ATCGACGAGGCGGCCAACGGCTGCGCCGCTGCACTCGAAGGACACCCGGATTACAAGTAC 420

Qy   421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGCCGGGCGGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGCCGGGCGGG 480

Qy   481 CTCCTGGCGCCGGGCGGCGGGGG 500
           |||||||||||||||||||||||
Db   481 CTCCTGGCGCCGGGCGGCGGGGG 500

RESULT 3
AL138691/c
LOCUS   AL138691   147665 bp   DNA   linear   PRI 16-JUN-2001
DEFINITION   Human DNA sequence from clone RP11-310D8 on chromosome 13, complete

BASE COUNT   275 a   578 c   536 g   213 t
ORIGIN

Query Match   100.0%;
Best Local Similarity   100.0%;   Pred. No. 7.2e-51;
Matches 500;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     1 ATGCTCCTGGACGCGGCGGGTCCCGAGTTCCGGCAGCTTCGGCGCCAC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGCTCCTGGACGCGGCGGGTCCCGAGTTCCGGCAGCTTCGGCGCCAC 60
```

RESULT 2
AX321400
LOCUS   AX321400   1602 bp   DNA   linear   PAT 15-DEC-2001
DEFINITION Sequence 417 from Patent WO0177168.
ACCESSION   AX321400
VERSION   AX321400.1 GI:17905325
KEYWORDS
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS   Iodes,M.J., Wang,T., Mohamath,R. and Indrias,C.Y.
TITLE   Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL   Patent: WO 0177168-A 417 18-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES   Location/Qualifiers
source   1..1602
                /organism="Homo sapiens"
                /db_xref="taxon:9606"

Query Match   100.0%;
Best Local Similarity   100.0%;   Pred. No. 7.2e-51;
Matches 500;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Db   1 ATGCTCCTGGACGCGGCGGGTCCCGAGTTCCGGCAGCTTCGGCGCCAC 60
Qy   61 CATCACCACTCCGCCGCGGCGGGTTCTCGGCGCGGCCATCGGGGTGACTG 120
Db   61 CATCACCACTCCGCCGCGGCGGGTTCTCGGCGCGGCCATCGGGGTGACTG 120
Qy   121 AGCCTGGCGGGCGGCGGCCAGAACGGGCTTCGTTGCGGGGAGGCCTTC 180
Db   121 AGCCTGGCGGGCGGCGGCCAGAACGGGCTTCGTTGCGGGGAGGCCTTC 180
Qy   181 AAGCTCAAACCCGGGCGGCCAGAGCTGTCCCCGGGCGTTCACGTGCGCAG 240
Db   181 AAGCTCAAACCCGGGCGGCCAGAGCTGTCCCCGGGCGTTCACGTGCGCAG 240
Qy   241 GGCCCCGGCGGCCTACCCCGGCGTTCCCGGTGCCCGCCAGCGCGTCGGGCC 300
Db   241 GGCCCCGGCGGCCTACCCCGGCGTTCCCGGTGCCCGCCAGCGCGTCGGGCC 300
Qy   301 CACGCGCGCGCACGTTGGCTCTCTACTCTCTGGGCGCCGCCCTTCAACTC 360
Db   301 CACGCGCGCGCACGTTGGCTCTCTACTCTCTGGGCGCCGCCCTTCAACTC 360
Qy   361 TTCCCGAGCGCGCGCTTCCGGGGGACTTCGGCGCGGGGCACGGGCTGTTC 420
Db   361 TTCCCGAGCGCGCGCTTCCGGGGGACTTCGGCGCGGGGCACGGGCTGTTC 420
Qy   421 GGGCCCGCGGGCGGCGGGCCGCTGCACCACGGCGCCACCTCCTCTTC 480
Db   421 GGGCCCGCGGGCGGCGGGCCGCTGCACCACGGCGCCACCTCCTCTTC 480
Qy   481 CCGGGGCCTGCCAGAGCAGCA 500
Db   481 CCGGGGCCTGCCAGAGCAGCA 500
```

RESULT 3
AF104902
LOCUS   AF104902   1602 bp   mRNA   linear   PRI 17-DEC-1998
DEFINITION Homo sapiens ZIC2 protein (ZIC2) mRNA, complete cds.
ACCESSION   AF104902
VERSION   AF104902.1 GI:4028591
KEYWORDS
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1602)
AUTHORS   Brown,S.A., Warburton,D., Brown,L.Y., Yu,C.Y., Roeder,E.R.,
            Stengel-Rutkowski,S., Hennekam,R.C. and Muenke,M.
TITLE   Holoprosencephaly due to mutations in ZIC2, a homologue of
            Drosophila odd-paired
JOURNAL   Nat. Genet. 20 (2), 180-183 (1998)
MEDLINE   98442655
REFERENCE   2 (bases 1 to 1602)
AUTHORS   Brown,S., Brown,L.Y. and Warburton,D.
TITLE   Direct Submission
JOURNAL   Submitted (06-NOV-1998) Obstetrics and Gynecology, Columbia
            University, 630 W. 168th St., New York, NY 10032, USA
FEATURES   Location/Qualifiers
source   1..1602
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="13q32"
                /chromosome="13"
gene   1..1602
                /gene="ZIC2"
CDS   1..1602
                /gene="ZIC2"
                /codon_start=1
                /product="ZIC2 protein"
                /protein_id="AAC96325.1"
                /db_xref="GI:4028592"

BASE COUNT   275 a   578 c   536 g   213 t
ORIGIN

Query Match   100.0%;
Best Local Similarity   100.0%;   Pred. No. 7.2e-51;
Matches 500;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     1 ATGCTCCTGGACGCGGCGGGTCCCGAGTTCCGGCAGCTTCGGCGCCAC 60
```

Left panel

```
RESULT 4
LOCUS       AF193855    2680 bp    mRNA    linear   PRI 05-DEC-2000
DEFINITION  Homo sapiens zinc finger protein of cerebellum zic2 (ZIC2) mRNA,
            complete cds.
ACCESSION   AF193855
VERSION     AF193855.1  GI:11065969
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2680)
  AUTHORS   Yang,Y.; Hwang,C.K.; Junn,E.; Lee,G. and Mouradian,M.M.
  TITLE     ZIC2 and Sp3 Repress Sp1-induced Activation of the Human D1A
            Dopamine Receptor Gene
  JOURNAL   J. Biol. Chem. 275 (49), 38863-38869 (2000)
  PUBMED    1098499
REFERENCE   2  (bases 1 to 2680)
```

```
BASE COUNT    275 a    578 c    536 g    213 t
ORIGIN
```

```
/translation="MLLDAGPQFPAIGVSFARHHHSAAAAAAAEMQDRELSLAA
AQNGFVDSAAAAHMGAFKLNPGAHELSPGQSSAFTSQGPGAYPGSAAAAAAAALGPHA
AHVGSYSGPPFNSTRDFLFRSARLPGTSAPGGQHGLFGPGAGGLHHAHSDAQGHLLF
PGLPEOHGPHGSQNVLNGQOMRLGLPGEVFGRSEOYRQVASPRTDPYSAQLHQYGPM
NNMGMWMAAAAHHHHHHHHHPGAFFRYMRQQCIKQELICKWIDPEQLSNPKKSCNK
TFSTMHELVTHVSVEHVGGPEOSNHVCFWEECPREGKPFKAKYKLVNHIRVHTGEKPF
PCPFPGQCGKVFARSENIKIHKRTHTGEKPFQCEFEGCDRRFANSSDRKHMVHTSDK
PYLCKMCDKSYTHPSSLRKHMKVHESSPQGSESSPAASSGYESSTPPGLVSPSAEPOS
SSNLSPAAAAAAAAAAAAVSAVHRGGSGSGGGSGGSGSGGGSGGGSGGGGSGGGGGG
GSSGGGSGTGHSGLSSNFNEWYV"
```

```
Query Match            100.0%;   Score 500;   DB 9;   Length 1602;
Best Local Similarity  100.0%;   Conservative 0;   Pred. No. 7.2e-51;
Matches 500;   Mismatches 0;   Indels 0;   Gaps 0;

Qy    1 ATGTCTCCTGGACGCGGGGTCGGCCAGTTCCCGGGGCCAGCTTCGCGCGCCAC   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGTCTCCTGGACGCGGGGTCGGCCAGTTCCCGGGGCCAGCTTCGCGCGCCAC   60

Qy   61 CATCACCACTCCGCGCGCGCGCGGGGCGCCCAGATGCAGGACCGTGAACTG   120
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATCACCACTCCGCGCGCGCGCGGGGCGCCCAGATGCAGGACCGTGAACTG   120

Qy  121 AGCCTGGCGGCCGCAGAACGGCCTTCGTTGATTCCGCGCCGCCGCGAGCCTTC  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AGCCTGGCGGCCGCAGAACGGCCTTCGTTGATTCCGCGCCGCCGCGAGCCTTC  180

Qy  181 AAGCTCAACCCGGGCGCGCACGAGCTGTGTCCCCGGGCCAGAGCTCACGTCGCAG  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAGCTCAACCCGGGCGCGCACGAGCTGTGTCCCCGGGCCAGAGCTCACGTCGCAG  240

Qy  241 GGCCCCCGGGCCTACCCGCGCTCCCCGGCTGCGCGGCCGCTCGGCGGGCCC    300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GGCCCCCGGGCCTACCCGCGCTCCCCGGCTGCGCGGCCGCTCGGCGGGCCC    300

Qy  301 CACGCGCGCGCCACGTTGGCTCCTCTACTCTGGGCGCGCGCCCTTCAACTCCACCGGGACTTCCTG  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CACGCGCGCGCCACGTTGGCTCCTCTACTCTGGGCGCGCGCCCTTCAACTCCACCGGGACTTCCTG  360

Qy  361 TTCCGCAGCGGCGGCGGCGGCTTCGGGGTTCGGGGCGGCGGCGGCGGCGGC   420
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTCCGCAGCGGCGGCGGCGGCTTCGGGGTTCGGGGCGGCGGCGGCGGCGGC   420

Qy  421 GGGCCCGGCGGGCCGGGGGGGGCCGCGGGGGCCACTCGGACACGGGCCTGGTTC  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GGGCCCGGCGGGCCGGGGGGGGCCGCGGGGGCCACTCGGACACGGGCCTGGTTC  480

Qy  481 CCGGGCCTGCCAGAGAGCAGCA  500
        ||||||||||||||||||||||
Db  481 CCGGGCCTGCCAGAGAGCAGCA  500
```

Right panel

```
  AUTHORS   Yang,Y. and Mouradian,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) NINDS, NIH, 10 Center Drive, Bethesda, MD
            20892, USA
FEATURES             Location/Qualifiers
     source          1..2680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1..2680
                     /gene="ZIC2"
     CDS             1..1599
                     /gene="ZIC2"
                     /note="transcription factor; involved in cerebellar
                     development"
                     /codon_start=1
                     /product="zinc finger protein of cerebellum ZIC2"
                     /protein_id="AAG28409.1"
                     /db_xref="GI:11065970"
                     /translation="MLLDAGPQFPAIGVSFARHHHSAAAAAAAEMQDRELSLAA
                     AQNGFVDSAAAAHMGAFKLNPGAHELSPGQSSAFTSQGPGAYPGSAAAAAAAALGPHA
                     AHVGSYSGPPFNSTRDFLFRSRGRGRSEOYRQVASPRTDPYSAQLHQYGPMN
                     GLPEOHGPHGSQNVLNGQOMRLGLPGEVFGRSEOYRQVASPRTDPYSAQLHQYGPMN
                     NMGMWMAAAAHHHHHHHHHPGAFFRYMRQQCIKQELICKWIDPEQLSNPKKSCNKT
                     FSTMHELVTHVSVEHVGGPEOSNHVCFWEECPREGKPFKAKYKLVNHIRVHTGEKP
                     FPCPFPGQCGKVFARSENIKIHKRTHTGEKPFQCEFEGCDRRFANSSDRKHMVHTSDKP
                     YLCKMCDKSYTHPSSLRKHMKVHESSPQGSESSPAASSGYESSTPPGLVSPSAEPOSS
                     SNLSPAAAAAAAAAAAAVSAVHRGGSGSGGGSGGSGSGGGSGGGSGGGGSGGGGGG
                     SSGGGSGTGHSGLSSNFNEWYV"
```

```
BASE COUNT    572 a    798 c    723 g    586 t    1 others
ORIGIN
```

```
Query Match            93.5%;   Score 467.6;   DB 9;   Length 2680;
Best Local Similarity  98.6%;   Pred. No. 4.2e-47;
Matches 493;   Conservative 0;   Mismatches 4;   Indels 3;   Gaps 2;

Qy    1 ATGTCTCCTGGACGCGGGGTCGGCCAGTTCCCGGGGCCATCGGGGGTGGGCAGCTTCGCGCGCCAC  60
        ||||||||||||||||||||||||||||||||||||||
Db    1 ATGTCTCCTGGACGCGGGGTCGGCCAGTTCCCGGGGCCATCGGGGGTGGGCAGCTTCGCGCGCCAC  60

Qy   61 CATCACCACTCCGCGCGCGCGCGGGGCGCCCAGATGCAGGACCGTGAACTG  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATCACCACTCCGCGCGCGCGCGGGGCGCCCAGATGCAGGACCGTGAACTG  120

Qy  121 AGCCTGGCGGCCGCAGAACGGCCTTCGTTGACTCCGCGCCGCGAGCCTTC  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AGCCTGGCGGCCGCAGAACGGCCTTCGTTGACTCCGCGCCGCGAGCCTTC  180

Qy  181 AAGCTCAACCCGGGCGCGCACGAGCTGTCCCCGGGTTCACGTCGCAG  240
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAGCTCAACCCGGGCGCGCACGAGCTGTCCCCGGGTTCACGTCGCAG  240

Qy  241 GGCCCCGCGGGCCTCACCCGCCTGCGGGCCGCTCGGGGGCCC  300
        |||||||||||||||||||||||||||||||||||||||||
Db  241 GGCCCCGCGGGCCTCACCCGCCTGCGGGCCGCTCGGGGGCCC  300

Qy  301 CACGCGCGCCACGTTGGCTCCTACTCTGGGCGCGCGCCCTTCAACTCCACCGGGACTTCCTG  360
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CACGCGCGCCACGTTGGCTCCTACTCTGGGCGCGCGCCCTTCAACTCCACCGGGACTTCCTG  360

Qy  361 TTCCGCAGC-CGCGGCTTC--GGGGACTTCGGGGGCGGCGGCGGCTGTTC  420
        ||||||||   |      ||           |||||||||||||||||||
Db  361 TTCCGCAGC-CGCGGCTTC--GGGGACTTCGGGGGCGGCGGCGGCTGTTC  417

Qy  421 GGGCCCGGCGGGCCGGGGGGCCGCGGGGGCCACTCGGACACGGGCCTGGTTC  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 GGGCCCGGCGGGCCGGGGGGCCGCGGGGGCCACTCGGACACGGGCCTGGTTC  477

Qy  481 CCGGGCCTGCCAGAGAGCAGCA  500
        ||||||||||||||||||||||
Db  478 CCGGGCCTGCCAGAGAGCAGCA  497
```

BASE COUNT   267 a   376 c   417 g   262 t
ORIGIN

Query Match
Best Local Similarity   100.0%;   Score 500; DB 6;   Length 1322;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GGAATTCCGGGCGCCGGGTTGTGAGTAGTAGTTACCGGGAGTGGGGTGATCCCGGGCTAGGGGAGC   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAATTCCGGGCGCCGGGTTGTGAGTAGTAGTTACCGGGAGTGGGGTGATCCCGGGCTAGGGGAGC   60

Qy     61 GCGGCGCCCGATCGGCGGCTTAGTCGGAGCTCCGAAGGGAGTGACTAGGACACCCGGGTGGG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGGCGCCCGATCGGCGGCTTAGTCGGAGCTCCGAAGGGAGTGACTAGGACACCCGGGTGGG  120

Qy    121 CTACTTTTCTTCCGGTGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTCGGGCGTCGACGTTCGCCC  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTACTTTTCTTCCGGTGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTCGGGCGTCGACGTTCGCCC  180

Qy    181 ACTGAGCAAAAGAATTCCCTCGTAAAAACCCAGAGGACCCTCCCCGTCCTCGG  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACTGAGCAAAAGAATTCCCTCGTAAAAACCCAGAGGACCCTCCCCGTCCTCGG  240

Qy    241 GAGTGTCGCGGTCCCCCAGCCGCCGCCGGCAAAGGGAGCCGAGCCCCGGCCGCG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGTGTCGCGGTCCCCCAGCCGCCGCCGGCAAAGGGAGCCGAGCCCCGGCCGCG  300

Qy    301 GACGGGGGCCCCGGAGCTTGCCTGCCTCCCTCGCCCTCGCCCGGGGTTCGCTGCGGTAGAG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGGGGCCCCGGAGCTTGCCTGCCTCCCTCGCCCTCGCCCGGGGTTCGCTGCGGTAGAG  360

Qy    361 CGGAGGGGGGCGCGCGATGAAGGCGGTGAGCCCCGGTTCCGCCCCTCGGGCCGCAAGGCGCG  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGGAGGGGGGCGCGCGATGAAGGCGGTGAGCCCCGGTTCCGCCCCTCGGGCCGCAAGGCGCG  420

Qy    421 TCGGGCTGCGGGCGCGGGGAGCTGCGGCCGGTGCCCGACCACGGCCCACAGCCTG  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGGGCTGCGGGCGCGGGGAGCTGCGGCCGGTGCCCGACCACGGCCCACAGCCTG  480

Qy    481 GGTGGGCTCCGCAGCCGCGGC  500
          |||||||||||||||||||||
Db    481 GGTGGGCTCCGCAGCCGCGGC  500
```

RESULT 2
LOCUS   HSU28368   1322 bp   mRNA   linear   PRI 03-AUG-1995
DEFINITION Human Id-related helix-loop-helix protein Id4 mRNA, complete cds.
ACCESSION   U28368
VERSION   U28368.1   GI:881545
KEYWORDS
SOURCE   human.
ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1322).
  AUTHORS   Riesling,T.L. and Christy,B.A.
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1322)
  AUTHORS   Riesling,T.L.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1995) Traci L. Kiesling, Institute of Biotech,
    UTHSCSA, 15355 Lambda Dr., San Antonio, TX 78245, USA
FEATURES   Location/Qualifiers
  source   1..1322
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /sex="female"
              /tissue_type="abdominal adipose tissue"
              /dev_stage="15-year-old"
              /note="Caucasian"
  CDS   376..861
              /note="DNA binding protein inhibitor; Id-related
              helix-loop-helix protein"
              /codon_start=1
              /product="Id4"
              /protein_id="AAA73923.1"
              /db_xref="GI:881546"
              /translation="MKAVSPVRPSGRKAPSGCGGGBLALRCLAEHGHSLGGSAAAAA
              AAAARCKAAEAAADEPALCLQCDMNDCYSRLRRLVPTIPPNKKVSKVEILQHVIDYIL
              DLQLALETHPALLRQPPPPAPPHHPAGTCPAAPPRTPLTALNTDPAGAVNKQGDSILC
              R"
  misc_feature   565..687
              /note="encodes helix-loop-helix region"

BASE COUNT   267 a   376 c   417 g   262 t
ORIGIN

Query Match
Best Local Similarity   100.0%;   Score 500; DB 9;   Length 1322;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GGAATTCCGGGCGCCGGGTTGTGAGTAGTAGTTACCGGGAGTGGGGTGATCCCGGGCTAGGGGAGC   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAATTCCGGGCGCCGGGTTGTGAGTAGTAGTTACCGGGAGTGGGGTGATCCCGGGCTAGGGGAGC   60

Qy     61 GCGGCGCCCCGATCGGCGGCTTAGTCGGAGCTCCGAAGGGAGTGACTAGGACACCCGGGTGGG  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGGCGCCCGATCGGCGGCTTAGTCGGAGCTCCGAAGGGAGTGACTAGGACACCCGGGTGGG  120

Qy    121 CTACTTTTCTTCCGGTGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTCGGGCGTCGACGTTCGCCC  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTACTTTTCTTCCGGTGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTCGGGCGTCGACGTTCGCCC  180

Qy    181 ACTGAGCAAAAGAATTCCCTCGTAAAAACCCAGAGGACCCTCCCCGTCCTCGG  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACTGAGCAAAAGAATTCCCTCGTAAAAACCCAGAGGACCCTCCCCGTCCTCGG  240

Qy    241 GAGTGTCGCGGTCCCCCAGCCGCCGCCGGCAAAGGGAGCCGAGCCCCGGCCGCG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGTGTCGCGGTCCCCCAGCCGCCGCCGGCAAAGGGAGCCGAGCCCCGGCCGCG  300

Qy    301 GACGGGGGCCCCGGAGCTTGCCTGCCTCCCTCGCCCTCGCCCGGGGTTCGCTGCGGTAGAG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGGGGCCCCGGAGCTTGCCTGCCTCCCTCGCCCTCGCCCGGGGTTCGCTGCGGTAGAG  360

Qy    361 CGGAGGGGGGCGCGCGATGAAGGCGGTGAGCCCCGGTTCCGCCCCTCGGGCCGCAAGGCGCG  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGGAGGGGGGCGCGCGATGAAGGCGGTGAGCCCCGGTTCCGCCCCTCGGGCCGCAAGGCGCG  420

Qy    421 TCGGGCTGCGGGCGCGGGGAGCTGCGGCCGGTGCCCGACCACGGCCCACAGCCTG  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGGGCTGCGGGCGCGGGGAGCTGCGGCCGGTGCCCGACCACGGCCCACAGCCTG  480

Qy    481 GGTGGGCTCCGCAGCCGCGGC  500
          |||||||||||||||||||||
Db    481 GGTGGGCTCCGCAGCCGCGGC  500
```

RESULT 3
LOCUS   BC014941   2389 bp   mRNA   linear   PRI 29-OCT-2001
DEFINITION Homo sapiens, inhibitor of DNA binding 4, dominant negative
    helix-loop-helix protein, clone MGC:20126 IMAGE:4552357, mRNA,
    complete cds.
ACCESSION   BC014941
VERSION   BC014941.1   GI:15928943
KEYWORDS   MGC.
SOURCE   human.
ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2389)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
    Gene Collection (MGC), Cancer Genomics Office, National Cancer

BASE COUNT   419 a   805 c   731 g   434 t
ORIGIN
Query Match
Best Local Similarity 100.0%;   Pred. No. 1.6e-45;   Length 2389;
Matches 500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy     1 CGGGCTCAGCGGGGGGCCGGAGGGCCATGTTCCCGGTGTTGTTCCTTGCACGCGTGCTGGCCCCCC 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGGGCTCAGCGGGGGGCCGGAGGGCCATGTTCCCGGTGTTGTTCCTTGCACGCGTGCTGGCCCCCC 60

Qy    61 CTTCCCGGTGCTGGGCCTGGACTCCCGGGGGGGGCGGGTGGGGGGCGGCCTCATGAACTCCTTCCCGCC 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTTCCCGGTGCTGGGCCTGGACTCCCGGGGGGGGCGGGTGGGGGGCGGCCTCATGAACTCCTTCCCGCC 120

Qy   121 ACCTCAGGGTGCACCCCCAGAGAACCCCCTGCCAGTCCAGGTCAGGTGCGGGGCTCCAGTCCCGCTTCTT 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACCTCAGGGTGCACCCCCAGAGAACCCCCTGCCAGTCCAGGTCAGGTGCGGGGCTCCAGTCCCGCTTCTT 180

Qy   181 TGCCTCTCCAGGGCTGCCCCCAGAGAGTCCATTCCAGGCGCGCGGGGCTGCCCCCCGCCCCACGCC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGCCTCTCCAGGGCTGCCCCCAGAGAGTCCATTCCAGGCGCGCGGGGCTGCCCCCCGCCCCACGCC 240

Qy   241 CCAGGCCCGGCGCGGGGCCGAGCCCCTCCAGGTGCTGCCGCCCGCGGGCCCCCGCCCCAGCCA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCAGGCCCGGCGCGGGGCCGAGCCCCTCCAGGTGCTGCCGCCCGCGGGCCCCCGCCCCAGCCA 300

Qy   301 GGAGTCCGGCCGCGGGCTGCGCCCGCCCCCTGCTGCCGCCCGTCGCTGCCGCGCCGCGC 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGAGTCCGGCCGCGGGCTGCGCCCGCCCCCTGCTGCCGCCCGTCGCTGCCGCGCCGCGC 360

Qy   361 CCGGCCCCCTGCGGGGCCCCTCTACGGTGGGACACAGCGGCCCGCTCCGGCCGGGC 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCGGCCCCCTGCGGGGCCCCTCTACGGTGGGACACAGCGGCCCGCTCCGGCCGGGC 420

Qy   421 CCCTCCGCGCACCCCCGCCAGTGTCGGCGGCCCTGAAGCAGCAGCCTTCCGGCGCC 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCCTCCGCGCACCCCCGCCAGTGTCGGCGGCCCTGAAGCAGCAGCCTTCCGGCGCC 480

Qy   481 CGCCACTATCGCGCGGCGCGG 500
         |||||||||||||||||||||
Db   481 CGCCACTATCGCGCGGCGCGG 500
```

RESULT 2
HUMMAZ
LOCUS   HUMMAZ   2389 bp   mRNA   linear   PRI 27-APR-1993
DEFINITION Human zinc finger protein (MAZ) mRNA.
ACCESSION   M94046
VERSION   M94046.1 GI:187393
KEYWORDS   zinc-finger protein.
SOURCE   Homo sapiens female cDNA to mRNA.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2389)
  AUTHORS   Bossone,S.A., Asselin,C., Patel,A.J. and Marcu,K.B.
  TITLE   MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
            regulating transcriptional initiation and termination
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456 (1992)
  MEDLINE   92366479
FEATURES   Location/Qualifiers
     source   1..2389
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /sex="female"
                    /cell_line="Hela"

BASE COUNT   419 a   805 c   731 g   434 t
ORIGIN
Query Match
Best Local Similarity 100.0%;   Score 500; DB 6; Length 2389;
Matches 500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy     1 CGGGCTCAGCGGGGGGCCGGAGGGCCATGTTCCCGGTGTTGTTCCTTGCACGCGTGCTGGCCCCCC 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGGGCTCAGCGGGGGGCCGGAGGGCCATGTTCCCGGTGTTGTTCCTTGCACGCGTGCTGGCCCCCC 60

Qy    61 CTTCCCGGTGCTGGGCCTGGACTCCCGGGGGGGGCGGGTGGGGGGCGGCCTCATGAACTCCTTCCCGCC 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTTCCCGGTGCTGGGCCTGGACTCCCGGGGGGGGCGGGTGGGGGGCGGCCTCATGAACTCCTTCCCGCC 120

Qy   121 ACCTCAGGGTGCACCCCCAGAGAACCCCCTGCCAGTCCAGGTCAGGTGCGGGGCTCCAGTCCCGCTTCTT 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACCTCAGGGTGCACCCCCAGAGAACCCCCTGCCAGTCCAGGTCAGGTGCGGGGCTCCAGTCCCGCTTCTT 180

Qy   181 TGCCTCTCCAGGGCTGCCCCCAGAGAGTCCATTCCAGGCGCGCGGGGCTGCCCCCCGCCCCACGCC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGCCTCTCCAGGGCTGCCCCCAGAGAGTCCATTCCAGGCGCGCGGGGCTGCCCCCCGCCCCACGCC 240

Qy   241 CCAGGCCCGGCGCGGGGCCGAGCCCCTCCAGGTGCTGCCGCCCGCGGGCCCCCGCCCCAGCCA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCAGGCCCGGCGCGGGGCCGAGCCCCTCCAGGTGCTGCCGCCCGCGGGCCCCCGCCCCAGCCA 300

Qy   301 GGAGTCCGGCCGCGGGCTGCGCCCGCCCCCTGCTGCCGCCCGTCGCTGCCGCGCCGCGC 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGAGTCCGGCCGCGGGCTGCGCCCGCCCCCTGCTGCCGCCCGTCGCTGCCGCGCCGCGC 360

Qy   361 CCGGCCCCCTGCGGGGCCCCTCTACGGTGGGACACAGCGGCCCGCTCCGGCCGGGC 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCGGCCCCCTGCGGGGCCCCTCTACGGTGGGACACAGCGGCCCGCTCCGGCCGGGC 420

Qy   421 CCCTCCGCGCACCCCCGCCAGTGTCGGCGGCCCTGAAGCAGCAGCCTTCCGGCGCC 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCCTCCGCGCACCCCCGCCAGTGTCGGCGGCCCTGAAGCAGCAGCCTTCCGGCGCC 480

Qy   481 CGCCACTATCGCGCGGCGCGG 500
         |||||||||||||||||||||
Db   481 CGCCACTATCGCGCGGCGCGG 500
```

RESULT 3
HUMZINCFP
LOCUS   HUMZINCFP   1638 bp   mRNA   linear   PRI 24-NOV-1994
DEFINITION Human zinc finger protein mRNA.
ACCESSION   M93339 J05371
VERSION   M93339.1 GI:340439
KEYWORDS   zinc-finger protein.
SOURCE   Homo sapiens (library: lambda gt11) female cDNA to mRNA.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1638)
  AUTHORS   Pyrc,J.J., Moberg,K.H. and Hall,D.J.
  TITLE   Isolation of a novel cDNA encoding a zinc-finger protein that binds
            to two sites within the c-myc promoter
  JOURNAL   Biochemistry 31 (16), 4102-4110 (1992)
  MEDLINE   92232709
FEATURES   Location/Qualifiers
     source   1..1638
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /sex="female"
                    /cell_line="Hela"
                    /cell_type="human cervical carcinoma, epithelial."
                    /tissue_lib="lambda gt11"

BASE COUNT   275 a   598 c   515 g   250 t
ORIGIN
Query Match
Best Local Similarity 98.5%;   Score 492.6; DB 9; Length 1638;
Matches 495; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

```
Qy     2 GGCTCAGCGGGGGGCGGGGGCCGGAGGGCCATGTTCCCGGTGTTGTTCCTTGCACGCGTGCTGGCCCCCCCC 61
         ||| || | ||||||||||   || ||||||||||||||||||||||||||||||||||||||||||||||||
Db    31 GGGCCCGGCCTTGGGCGGCCGAGGGCCATGTTCCCGGGTGTTGTTCCTTGCACGCGTGCTGGCCCCCCCC 90
```

RESULT 2

```
BASE COUNT     738 a    312 c    413 g    397 t
ORIGIN

Query Match
Best Local Similarity  100.0%;  Score 500; DB 6; Length 1860;
Matches 500;  Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    1 GGGACGTGAGCCGCTGCGCCCGCTAGACCCGGCCAATCATGCTGCTTCTGCCAA  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGGACGTGAGCCGCTGCGCCCGCTAGACCCGGCCAATCATGCTGCTTCTGCCAA  60

Qy   61 GCGCGCCGGACGCGGCGGGCGGGCACCGCCATCACCCACGCTCTCGACCTCTGC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCGCGCCGGACGCGGCGGGCGGGCACCGCCATCACCCACGCTCTCGACCTCTGC 120

Qy  121 GTCAAGTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTTAAGAGAGAAACAGAAATG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GTCAAGTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTTAAGAGAGAAACAGAAATG 180

Qy  181 CTTCTGCCTCTTTTCAGGAACTGGAGAGTTTCCCATGCTGAAAAACACTTACACTCT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTTCTGCCTCTTTTCAGGAACTGGAGAGTTTCCCATGCTGAAAAACACTTACACTCT 240

Qy  241 AAGAATTGCAGTTGGGAAGAGTTTCCCATGCTGAAAAACACTTGACACAGAGAC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAGAATTGCAGTTGGGAAGAGTTTCCCATGCTGAAAAACACTTGACACAGAGAC 300

Qy  301 ACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGATACAAGGCTACACAGAGAC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGATACAAGGCTACACAGAGAC 360

Qy  361 AGATCAAAGCAGCTCAATAAAGCAATGGTTTAAAACATCACCCAGACAAACGGAAAGCAG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AGATCAAAGCAGCTCAATAAAGCAATGGTTTAAAACATCACCCAGACAAACGGAAAGCAG 420

Qy  421 CTGGTGAACCAATAAAAGAAGGAGAATAATGACTACTTCACTTGCATAACTAAAGCTTATG 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 CTGGTGAACCAATAAAAGAAGGAGAATAATGACTACTTCACTTGCATAACTAAAGCTTATG 480

Qy  481 AAATGTTATCTGATCCAGTG 500
        ||||||||||||||||||||
Db  481 AAATGTTATCTGATCCAGTG 500
```

```
LOCUS       HSMPP11               1860 bp    mRNA    linear   PRI 08-JAN-1997
DEFINITION  H.sapiens mRNA for M-phase phosphoprotein, mpp11.
ACCESSION   X98260
VERSION     X98260.1  GI:1770453
KEYWORDS    M phase phosphoprotein; MPP gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1860)
  AUTHORS   Matsumoto-Taniura,N., Pirollet,F., Monroe,R., Gerace,L. and
            Westendorf,J.M.
  TITLE     Identification of novel M phase phosphoproteins by expression
            cloning
  JOURNAL   Mol. Biol. Cell 7 (9), 1455-1469 (1996)
  MEDLINE   97039687
REFERENCE   2  (bases 1 to 1860)
  AUTHORS   Westendorf,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1996) J.M. Westendorf, INSERM U366, DBMS/CS-CENG,
            17 rue des Martyrs, F-38054 Grenoble Cedex 9, FRANCE
FEATURES             Location/Qualifiers
     source          1..1860
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="14"
                     /sex="male"
                     /cell_line="MOLT-4"
                     /cell_type="lymphoblast-like"
                     /tissue_type="blood"
                     /clone_lib="lambda gt11"
                     /dev_stage="19 years old"
     gene            1..1751
                     /gene="mpp11"
     CDS             1..1751
                     /gene="mpp11"
                     /codon_start=3
                     /product="M-phase phosphoprotein 11"
                     /protein_id="CAA66913.1"
                     /db_xref="GI:1770454"
                     /db_xref="SPTREMBL:Q99543"
                     /translation="DVSRCAHRARPGAIMLLPSAADGRGTAITHALTSASTLCQVEP
                     VGRWFEAFVKRNRNASASFQELEDKKELSESEDEELQLEEFPMLKILDPKDWKNQD
                     HYAVLGLGHVRYKATORQIKAAHKAMYLKHHPDRKAAGEPIKEGDNDFTCIYKAYE
                     MLSDPVKRRAFNSVDFTFDNSYPSKSEAKDNFEVFPVFERNSRWSKNKKNVKLGDM
                     NSSFEDVDIFYSFWYNFDSWREFSYLDEEEKEKAECRDERRWIEKQNGATRAQRKKEE
                     MRRIRILVDNAYSCDPRIKKFKEEKAKAEAKRKEQEAKEKQROAELEAA
                     RLAKEKEEEVRQQALLAKKEDIQKAIKKEROKLRNSCKIEEINEQIKREKEAEA
                     RMRQASKNTEKSTGGGNGSKNWSEDDLQLLIKAVNLFPARTNSRWEVIANYMNIHSS
                     SGVRTAKDVIGRAKSLQKLDPHQKDDINKKAFDFKKEHGVVPQADNATPSERFEGP
                     YTDFTPWTTEEQKLLIEQALRTYPVNTEPERWEKIAEAVPGRTKKDCMKRYKELYEMVKA
                     KKAAQEQVLNASRAKK"
     polyA_signal    1779..1784
     polyA_site      1845
```

```
BASE COUNT     738 a    312 c    413 g    397 t
ORIGIN

Query Match
Best Local Similarity  100.0%;  Score 500; DB 9; Length 1860;
Matches 500;  Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    1 GGGACGTGAGCCGCTGCGCCCGCTAGACCCGGCCAATCATGCTGCTTCTGCCAA  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGGACGTGAGCCGCTGCGCCCGCTAGACCCGGCCAATCATGCTGCTTCTGCCAA  60

Qy   61 GCGCGCCGGACGCGGCGGGCGGGCACCGCCATCACCCACGCTCTCGACCTCTGC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCGCGCCGGACGCGGCGGGCGGGCACCGCCATCACCCACGCTCTCGACCTCTGC 120

Qy  121 GTCAAGTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTTAAGAGAGAAACAGAAATG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GTCAAGTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTTAAGAGAGAAACAGAAATG 180

Qy  181 CTTCTGCCTCTTTTCAGGAACTGGAGAGTTTCCCATGCTGAAAAACACTTACACTCT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTTCTGCCTCTTTTCAGGAACTGGAGAGTTTCCCATGCTGAAAAACACTTACACTCT 240

Qy  241 AAGAATTGCAGTTGGGAAGAGTTTCCCATGCTGAAAAACACTTGACACAGAGAC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAGAATTGCAGTTGGGAAGAGTTTCCCATGCTGAAAAACACTTGACACAGAGAC 300

Qy  301 ACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGATACAAGGCTACACAGAGAC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGATACAAGGCTACACAGAGAC 360

Qy  361 AGATCAAAGCAGCTCAATAAAGCAATGGTTTAAAACATCACCCAGACAAACGGAAAGCAG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AGATCAAAGCAGCTCAATAAAGCAATGGTTTAAAACATCACCCAGACAAACGGAAAGCAG 420

Qy  421 CTGGTGAACCAATAAAAGAAGGAGAATAATGACTACTTCACTTGCATAACTAAAGCTTATG 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 CTGGTGAACCAATAAAAGAAGGAGAATAATGACTACTTCACTTGCATAACTAAAGCTTATG 480

Qy  481 AAATGTTATCTGATCCAGTG 500
        ||||||||||||||||||||
Db  481 AAATGTTATCTGATCCAGTG 500
```

RESULT 3

```
MUSMIDA1
```

RESULT 2

```
HSU23028
LOCUS       HSU23028        2291 bp    mRNA      linear    PRI 22-AUG-1996
DEFINITION  Human eukaryotic initiation factor 2B-epsilon mRNA, partial cds.
ACCESSION   U23028
VERSION     U23028.1   GI:806853
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2291)
  AUTHORS   Asuru,A.I., Mellor,H., Thomas,N.S., Yu,L., Chen,J.J., Crosby,J.S.,
            Hartson,S.D., Kimball,S.R., Jefferson,L.S. and Matts,R.L.
  TITLE     Cloning and characterization of cDNAs encoding the epsilon-subunit
            of eukaryotic initiation factor-2B from rabbit and human
  JOURNAL   Biochim. Biophys. Acta 1307 (3), 309-317 (1996)
  MEDLINE   96305354
REFERENCE   2  (bases 1 to 2291)
  AUTHORS   Asuru,A.I., Mellor,H., Thomas,N.S.B., Yu,L., Chen,J.-J.,
            Crosby,J.S., Hartson,S.D., Kimball,S.R., Jefferson,L.S. and
            Matts,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1995) Robert L. Matts, Biochemistry & Molecular
            Biology, Oklahoma State University, 246 NRC, Stillwater, OK
            74078-0454, USA
FEATURES             Location/Qualifiers
     source          1..2291
                     /organism="Homo sapiens"
```

```
Query Match         100.0%;  Score 500;  DB 6;  Length 2291;
Best Local Similarity  100.0%;  Pred. No. 5.6e-143;
Matches   500;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  GAATTCCTGACTCGCCACAGGTGTACAGAGAAACAATTTGTCTTTGTTTGTTGCTGGAAAGCTTGCT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GAATTCCTGACTCGCCACAGGTGTACAGAGAAACAATTTGTCTTTGTTTGTTGCTGGAAAGCTTGCT  60

Qy     61  CAAATCAAAGAACAATTTACTGAAGTCAAAAGTGGTGCCGCCCTACATCTCTCAAGTGTGGTT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CAAATCAAAGAACAATTTACTGAAGTCAAAAGTGGTGCCGCCCTACATCTCTCAAGTGTGGTT  120

Qy    121  CGAATAATTACATCAGAGCTCTATCGATCACTCGGGAGATGTCCTCCGTGATGTGTGATGCC  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CGAATAATTACATCAGAGCTCTATCGATCACTCGGGAGATGTCCTCCGTGATGTGTGATGCC  180

Qy    181  AAGGCTTTGGTTGCGCTCTGACTTTTCTTCTCTTGGTTGTATGGGGATGTCATCTCAAACATCAAT  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  AAGGCTTTGGTTGCGCTCTGACTTTTCTTCTCTTGGTTGTATGGGGATGTCATCTCAAACATCAAT  240

Qy    241  ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAGCTAGAAAAAAAATGTTTCTGTG  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAGCTAGAAAAAAAATGTTTCTGTG  300

Qy    301  ATGACGATGATCTTCAAGGAGTCATCCCCAGCCACACCAACTCGTTGCCACGAAGACAAT  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||+||||||
Db    301  ATGACGATGATCTTCAAGGAGTCATCCCCAGCCACACCAACTCGTTGCCACGAAGACAAT  360

Qy    361  GTGGTAGTGGGCTGTGGATAGTAGTACCACAAAACAGGGTTCTCCATTTTCAGAAGACCCCAGGGT  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GTGGTAGTGGGCTGTGGATAGTAGTACCACAAAACAGGGTTCTCCATTTTCAGAAGACCCCAGGGT  420

Qy    421  CTCCGGCGTTTTGCATTTCCTCTGAGCCTGTTGTCAGGGCAGTAGTGATGAGTGGGAGGTT  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTCCGGCGTTTTGCATTTCCTCTGAGCCTGTTGTCAGGGCAGTAGTGATGAGTGGGAGGTT  480

Qy    481  CGATATGAGTTTACTGGATTG  500
           |||||||||||||||||||||
Db    481  CGATATGAGTTTACTGGATTG  500
```

```
     CDS             <1..1926
                     /codon_start=1
                     /note="epsilon subunit of eukaryotic initiation factor 2B"
                     /product="eIF-2Bepsilon"
                     /protein_id="AAC50646.1"
                     /db_xref="GI:806854"
                     /translation="EFLTATGVQETFVFCCWKAAQIKEHLLKSKWCRPSLNVRIIT
                     SELYRSLGDVLRDVDAKALVRSDFLLVIGDVISNINITRALEEHRLRKLEKNVSVMT
                     MIFKESSPSHPTRCHEDNVVAVDSTTNRVLHFQKTQGLRRFAFPLSLFQGSSDGVEV
                     RYDLLDCHISISPQVAQLFTDNFDYQTRDDFVRGLLVNEEILGNQIHHVTAKEYGA
                     RVSNLHMXSAVCADVIRRWYPLTPEANFTDSTTQOSCTHSRHNIYRGPEVSLGHOSIL
                     EENVLLGSGTVIGSNCFITNSVIGPGCHIGDNVVLDQTYLMQGYRVAAGAQIHOSLLC
                     DNAEVKRVTLKPRSVLTSQVVVGPNIILPEGSVISLSHPPDAEEDEDGEFSDSGAD
                     QEKDKVKMKGVNPAEVGRAGKGYIWKAAGMMEEEEELQNLWGLKIMEEESESESE
                     QSMDSEEPDSRGGSPQMDDIKVFQNEVLGTLQRSKEENTSCDNLVLEINSLKYAVVS
                     LKEVMQVLSHVVLEFPLQQOMDSPLDSSSRYCALLLPLLKAWSPVFRNYIKRAADHLEAL
                     AAIEDFPLEHEALGISMAKVLMAPYQLEIILAEEFIILSWFSQRDTTDKGQOLRKNOOLQ
                     RFIQWLKRAEEESSEDD"

     polyA_site      2291
                     /note="21 A nucleotides"
```

RESULT 3

```
BASE COUNT        588 a        540 c        630 g        533 t
ORIGIN
```

```
Query Match         100.0%;  Score 500;  DB 9;  Length 2291;
Best Local Similarity  100.0%;  Pred. No. 5.6e-143;
Matches   500;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  GAATTCCTGACTCGCCACAGGTGTACAGAGAAACAATTTGTCTTTGTTTGTTGCTGGAAAGCTTGCT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GAATTCCTGACTCGCCACAGGTGTACAGAGAAACAATTTGTCTTTGTTTGTTGCTGGAAAGCTTGCT  60

Qy     61  CAAATCAAAGAACAATTTACTGAAGTCAAAAGTGGTGCCGCCCTACATCTCTCAAGTGTGGTT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CAAATCAAAGAACAATTTACTGAAGTCAAAAGTGGTGCCGCCCTACATCTCTCAAGTGTGGTT  120

Qy    121  CGAATAATTACATCAGAGCTCTATCGATCACTCGGGAGATGTCCTCCGTGATGTGTGATGCC  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CGAATAATTACATCAGAGCTCTATCGATCACTCGGGAGATGTCCTCCGTGATGTGTGATGCC  180

Qy    181  AAGGCTTTGGTTGCGCTCTGACTTTTCTTCTCTTGGTTGTATGGGGATGTCATCTCAAACATCAAT  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  AAGGCTTTGGTTGCGCTCTGACTTTTCTTCTCTTGGTTGTATGGGGATGTCATCTCAAACATCAAT  240

Qy    241  ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAGCTAGAAAAAAAATGTTTCTGTG  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAGCTAGAAAAAAAATGTTTCTGTG  300

Qy    301  ATGACGATGATCTTCAAGGAGTCATCCCCAGCCACACCAACTCGTTGCCACGAAGACAAT  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  ATGACGATGATCTTCAAGGAGTCATCCCCAGCCACACCAACTCGTTGCCACGAAGACAAT  360

Qy    361  GTGGTAGTGGGCTGTGGATAGTAGTACCACAAAACAGGGTTCTCCATTTTCAGAAGACCCCAGGGT  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GTGGTAGTGGGCTGTGGATAGTAGTACCACAAAACAGGGTTCTCCATTTTCAGAAGACCCCAGGGT  420

Qy    421  CTCCGGCGTTTTGCATTTCCTCTGAGCCTGTTGTCAGGGCAGTAGTGATGAGTGGGAGGTT  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTCCGGCGTTTTGCATTTCCTCTGAGCCTGTTGTCAGGGCAGTAGTGATGAGTGGGAGGTT  480

Qy    481  CGATATGAGTTTACTGGATTG  500
           |||||||||||||||||||||
Db    481  CGATATGAGTTTACTGGATTG  500
```

```
BC013590
LOCUS       BC013590        2651 bp    mRNA      linear    PRI 07-SEP-2001
DEFINITION  Homo sapiens, clone MGC:9947 IMAGE:3876105, mRNA, complete cds.
ACCESSION   BC013590
VERSION     BC013590.1   GI:15488925
KEYWORDS    MGC.
```

```
BASE COUNT      481 a    352 c    359 g    388 t
ORIGIN
```

```
LOCUS       HUMRBPJKA     1580 bp    mRNA    linear    PRI 29-JAN-2001
DEFINITION  Homo sapiens recombination signal binding protein (RBPJK) mRNA,
            partial cds.
ACCESSION   L07872
VERSION     L07872.1  GI:190049
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1580)
  AUTHORS   Amakawa,R., Jing,W., Ozawa,K., Matsunami,N., Hanaguchi,Y.,
            Matsuda,F., Kawaichi,M. and Honjo,F.
  TITLE     Human JK recombination signal binding protein gene (IGKJRB):
            comparison with its mouse homologue
  JOURNAL   Genomics 17 (2), 306-315 (1993)
  MEDLINE   94010923
   PUBMED   8406481
FEATURES             Location/Qualifiers
     source          1..1580
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3q25"
                     /dev_stage="adult"
     gene            1..>1580
                     /gene="RBPJK"
     CDS             79..>1580
                     /gene="RBPJK"
                     /codon_start=1
                     /product="recombination signal binding protein"
                     /protein_id="AAA60258.1"
                     /db_xref="GI:553639"
                     /translation="MDHTEGLPAEEPPAHAPSPGKEFGERPPPKRLTREAMRNYLKERG
                     DQTVLILHAKVAQKSYGNEKRFFCPPPCVILMGSGWKKKKEQMERDGCSEQSSQPCAF
                     IGIGNSDQEMQQLNLEGKNYCTAKTLYISDSDKRKIFFSVKMFYGNSDDIGVFLSKR
                     IKVISKPSKKKQSLKNADLCIASGTRVALFNRLRSQTVSTRYLHVEGGNFHASSQWG
                     AFFIHLLDDDESEGEEFTVRDVYIHYGQTCKLVCSVTGMALPRLIIMKVDKHTALLDA
                     DDPVSQLHKCAFYLKDTERMLCLSQERIIQFQATPCPKEENKEMINDGASWTIISTD
                     KAEYTFYEGMGPVLAPVTPVPVVESLQLNGGGDVAMLELTGQNFTPNLRVWFGDVEAE
                     TMYRCGESMLCVVPDISAFREGWRWVRQPVQVPVTLVRNDGIIYSTSLTFTYTPEPGP
                     RPHCSVAGAILPANSSGGVPPNESNTNSEGSYTNASTNSTSVTSSSTATVVS"
     exon            1..137
                     /gene="RBPJK"
                     /number=1
     exon            138..176
                     /gene="RBPJK"
                     /number=2
     exon            177..272
                     /gene="RBPJK"
                     /number=3
     exon            273..438
                     /gene="RBPJK"
                     /number=4
     exon            439..613
                     /gene="RBPJK"
                     /number=5
     exon            614..751
                     /gene="RBPJK"
                     /number=6
     exon            752..864
                     /gene="RBPJK"
                     /number=7
     exon            865..1005
                     /gene="RBPJK"
                     /number=8
     exon            1006..1161
                     /gene="RBPJK"
                     /number=9
     exon            1162..1265
                     /gene="RBPJK"
                     /number=10
     exon            1266..1503
                     /gene="RBPJK"
                     /number=11
```

```
RESULT 2
Query Match             100.0%;  Score 500;  DB 6;  Length 1580;
Best Local Similarity   100.0%;  Pred. No. 4.2e-133;
Matches  500;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  ATCCCTCCGGTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGGCGTCCTRAAACCCGG   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      1  ATCCCTCCGGTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGGCGTCCTRAAACCCGG   60

Qy     61  ATAACCGGAGCGCGCTCCCCATGGACCACACGGAGGGCTTGCCCGCGCGGAGGAGCCGCCGCTGCG  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ATAACCGGAGCGCGCTCCCCATGGACCACACGGAGGGCTTGCCCGCGCGGAGGAGCCGCCGCTGCG  120

Qy    121  CATGCTCCATCGCCTGGGAAATTTGGTGAGCGGCGCCTCCACCTAAACGACTTACTAGGGAA  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CATGCTCCATCGCCTGGGAAATTTGGTGAGCGGCGCCTCCACCTAAACGACTTACTAGGGAA  180

Qy    181  GCTATGCGAAATTATTTAAAAGAGCCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAA  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GCTATGCGAAATTATTTAAAAGAGCCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAA  240

Qy    241  GTTGCACAGAAGTCATATGGAAATGAAAAAGGTTTTTTTTCGCCACCTCCTTGTGTATAT  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GTTGCACAGAAGTCATATGGAAATGAAAAAGGTTTTTTTTCGCCACCTCCTTGTGTATAT  300

Qy    301  CTTATGGGGCAGGCGGATGGAGAGAAAAGAAGAACAAAATGAAACGGCGATGGTTGTTCTGAA  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CTTATGGGGCAGGCGGATGGAGAGAAAAGAAGAACAAAATGAAACGGCGATGGTTGTTCTGAA  360

Qy    361  CAAGAGTCTCAACCGTGTGCAtTTTATTGGGATAGGAGAaATGACCAAGAAATGCAGCAG  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  CAAGAGTCTCAACCGTGTGCAtTTTATTGGGATAGGAGAaATGACCAAGAAATGCAGCAG  420

Qy    421  CTAAACTTGGAAGGAAAGAAACTATTGCACAGCCAAAACATTGTATAATATCTGACTCAGAC  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTAAACTTGGAAGGAAAGAAACTATTGCACAGCCAAAACATTGTATAATATCTGACTCAGAC  480

Qy    481  AAGGGAAAGCACACTTCATTTT   500
           ||||||||||||||||||||||
Db    481  AAGGGAAAGCACACTTCATTTT   500
```

```
BASE COUNT      481 a    352 c    359 g    388 t
ORIGIN
Query Match             100.0%;  Score 500;  DB 9;  Length 1580;
Best Local Similarity   100.0%;  Pred. No. 4.2e-133;
Matches  500;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  ATCCCTCCGGTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGGCGTCCTCTAAAACCCGG   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  ATCCCTCCGGTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGGCGTCCTCTAAAACCCGG   60

Qy     61  ATAACCGGAGCGCGCTCCCCATGGACCACACGGAGGGCTTGCCCGCGCGGAGGAGCCGCCGCTGCG  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ATAACCGGAGCGCGCTCCCCATGGACCACACGGAGGGCTTGCCCGCGCGGAGGAGCCGCCGCTGCG  120

Qy    121  CATGCTCCATCGCCTGGGAAATTTGGTGAGCGGCGCCTCCACCTAAACGACTTACTAGGGAA  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CATGCTCCATCGCCTGGGAAATTTGGTGAGCGGCGCCTCCACCTAAACGACTTACTAGGGAA  180

Qy    181  GCTATGCGAAATTATTTAAAAGAGCCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAA  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GCTATGCGAAATTATTTAAAAGAGCCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAA  240

Qy    241  GTTGCACAGAAGTCATATGGAAATGAAAAAGGTTTTTTTTCGCCCACCTCCTTGTGTATAT  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GCTATGCGAAATTATTTAAAAGAGCCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAA  240
```

RESULT 2

LOCUS   HSSOX3   2509 bp   DNA   linear   PRI 11-AUG-1994
DEFINITION H.sapiens sox3 gene.
ACCESSION   X71135
VERSION   X71135.1 GI:468790
KEYWORDS   sox3 gene.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2508)
  AUTHORS   Stevanovic.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-APR-1993) M. Stevanovic, University of Cambridge,
            Genetics Dept, Downing Street, Cambridge, CB2 3EH, UK
REFERENCE   2 (bases 1 to 2509)
  AUTHORS   Stevanovic.M., Lovell-Badge,R., Collignon,J. and Goodfellow,P.N.
  TITLE   SOX3 is an X-linked gene related to SRY
  JOURNAL   Hum. Mol. Genet. 2 (12), 2013-2018 (1993)
  MEDLINE   94154672
FEATURES   Location/Qualifiers
     source   1..2509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene   <1..>2508
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
     TATA_signal   404..407
                     /map="Xq26-Xq27"
                     /gene="sox3"
     gene   442..1773
                     /gene="sox3"
     CDS   442..1773
                     /gene="sox3"
                     /codon_start=1
                     /product=SOX3"
                     /protein_id="CAA50465.1"
                     /db_xref="GI:530020"
                     /db_xref="SWISS-PROT:P41225"
                     /translation="MRPVRENSSGARSPRVPADLARSILISLPFFPPDSLAHRPPSSAP
                     TESQGLFTVAAPAPGAPSPPATLAHLLPAPAMYSLLETELKNPVGTPTQAAGTGGPAA
                     PGGAGKSSANAAGGANSGGSGGSSGGASGGGGTQDDRVKRPMNAFMVWSRGQRRKMALE
                     NPKMHNSEISKRLGADWKLLTDAEKRPPIDEAKRLRAVHMKEYPDYKYRPRRKTKTLL
                     KKDKYSLPSGLLPPGAAAAAAAAAAASPVGVGQRLDTYTHVNGWANGAYSLVQ
                     EQLGYAQPSMSSPPPPPALHRYDMAGLQIYSPMMPGAQSYMNVAAAAAAASGYGGMA
                     PSATAAAAAYGQPATAAAAAAMSLGPMGSVVKSESPSSPPAIASHSQRACLG
                     DLRDMISMYLPPGDADAASPLPGGRLHGVHQHYQGAGTAVNGTVPLTHI"
                     /polyA_signal 2493..2498
BASE COUNT   456 a   818 c   757 g   478 t
ORIGIN

Query Match   100.0%; Score 500; DB 9; Length 2509;
Best Local Similarity   100.0%; Pred. No. 4.5e-135;
Matches   500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TGGCCGGGGGGATGGGGCGGCGGCCGGTCGCCTTGACAGGGTTGCAAAGTTGTTTCTAAATTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TGGCCGGGGGGATGGGGCGGCGGCCGGTCGCCTTGACAGGGTTGCAAAGTTGTTTCTAAATTC 60

Qy   61 CGAAGCGCCCCTCTGCCCCCCCAATCTGCTGCTGCGGGGGTGGGGGGTGGGGGGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CGAAGCGCCCCTCTGCCCCCCCAATCTGCTGCTGCGGGGGTGGGGGGTGGGGGGG 120

Qy   121 GTCACCTCCTCAGGTTTCGTTCTTCTTTCAAACATTTTTGAAAACCTAATTGGTTGGCCTCTGAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTCACCTCCTCAGGTTTCGTTCTTCTTTCAAACATTTTTGAAAACCTAATTGGTTGGCCTCTGAG 180

Qy   181 TGGGGCCTTCGTGGACTTCCCGCCCTCCTTAAGTAAACTCTTACCCACGTCACTAGGCCAAAGAGGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGGGCCTTCGTGGACTTCCCGCCCTCCTTAAGTAAACTCTTACCCACGTCACTAGGCCAAAGAGGG 240

Qy   241 GCGTGGGGTGAACGAAAAGGGCTCCCGAACTTTTTTTTTTTTTTTCCAGCCAGGCCGAACCGGGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCGTGGGGTGAACGAAAAGGGCTCCCGAACTTTTTTTTTTTTTTTCCAGCCAGGCCGAACCGGGGC 300

Qy   301 TCGGTAATGATTGGCCAGGGGCGCCATCACTGCGAACCTGTCAATCACGGGGTCCTCCGGGTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TCGGTAATGATTGGCCAGGGGCGCCATCACTGCGAACCTGTCAATCACGGGGTCCTCCGGGTT 360

Qy   361 GCGAGGGGGCGGACCAAGCCCCAACCCCGGGGAAATCCGACCAGGTATATATAAGGGGCCCAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GCGAGGGGGCGGACCAAGCCCCAACCCCGGGGAAATCCGACCAGGTATATATAAGGGGCCCAGC 420

Qy   421 TAGAGCCCAGGCAGACTGTGAATGCGACCTGTTCGAGAGAGAACTCATCAGGTGCCGGAGAAGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TAGAGCCCAGGCAGACTGTGAATGCGACCTGTTCGAGAGAGAACTCATCAGGTGCCGGAGAAGC 480

Qy   481 CCGCGGGGTTCCTGCTGATTT 500
          |||||||||||||||||||||
Db   481 CCGCGGGGTTCCTGCTGATTT 500

RESULT 3

LOCUS   HSBA51C14/c
DEFINITION Human DNA sequence from clone RP11-51C14 on chromosome Xq26.2-27.3,
            complete sequence.
ACCESSION   AL121875
VERSION   AL121875.10 GI:17055932
KEYWORDS   HTG.
SOURCE   human.

RESULT 2

BASE COUNT   1996 a   2143 c   2104 g   2124 t   5 others
ORIGIN

Query Match   100.0%; Score 500; DB 6; Length 8372;
Best Local Similarity   100.0%; Pred. No. 1.8e-144;
Matches 500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 AAGCTTGGTGCCATCTATTTTGGACTATGCCTTGCATACAGCTTTATGGGAACATTTGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTGGTGCCATCTATTTTGGACTATGCCTTGCATACAGCTTTATGGGAACATTTGTC 60

Qy     61 AGGCAAAAGTATATAATAAATGGCAAAACTCTACGCCCTTTTATTTTAAATTAGATTGGTGTGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGCAAAAGTATATAATAAATGGCAAAACTCTACGCCCTTTTATTTTAAATTAGATTGGTGTGAT 120

Qy    121 TTGATGCTGACGGGAGTGAGAGAGTAATGGCCCTTATCCTCGCGTGTGCTGAGGATGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGATGCTGACGGGAGTGAGAGAGTAATGGCCCTTATCCTCGCGTGTGCTGAGGATGG 180

Qy    181 CCTGGGTCTGCCACCCTCCTCGAGTAGCANTTTGCAGGGGGTCTCCCCTCTGGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTGGGTCTGCCACCCTCCTCGAGTAGCATTTTGCATGTGTAACAGGGGGTCTCCCCTCTGGG 240

Qy    241 GCACAACAACAAAAGAGAGAAGTTGCTAAGGACAAGAAGCAGGTGCGGGAAAATGCATCTCCCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCACAACAACAAAAGAGAGAAGTTGCTAAGGACAAGAAGCAGGTGCGGGAAAATGCATCTCCCAT 300

Qy    301 TGGAACAGACGCCTGGGCTTACTCCACATGGGCTGAGAGAGTGCTATGGCCAGTCCTCCCAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGAACAGACGCCTGGGCTTACTCCACATGGGCTGAGAGAGTGCTATGGCCAGTCCTCCCAGA 360

Qy    361 GCTCTCGAGCTGCACTTGGGGGTGGGACAGTTGGGGGTTTCCCATGGGCGGCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTCTCGAGCTGCACTTGGGGGTGGGACAGTTGGGGGTTTCCCATGGGCGGCA 420

Qy    421 GAAAGCCAGCCAACTGCTGCCCAAAAATCACCCAGCCGATTGGGGGTTTCCCATGGGCGGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAAGCCAGCCAACTGCTGCCCAAAAATCACCCAGCCGATTGGGGGTTTCCCATGGGCGGCA 480

Qy    481 CCCTGCCCGGAGCCAAGAAG 500
          ||||||||||||||||||||
Db    481 CCCTGCCCGGAGCCAAGAAG 500
```

LOCUS   AF107044   8372 bp   DNA   linear   PRI 13-DEC-1998
DEFINITION   Homo sapiens clone pCL4 DNA-binding protein SOX21 (SOX21) gene,
             complete cds.
ACCESSION   AF107044
VERSION   AF107044.1   GI:4008102
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 8372)
  AUTHORS   Malas,S., Duthie,S. and Episkopou,V.
  TITLE   The cloning and chromosomal localization of human SOX14 and SOX21,
             two members of the SOX gene family related to SOX1, SOX2 and SOX3
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 8372)
  AUTHORS   Malas,S., Duthie,S. and Episkopou,V.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-NOV-1998) Clinical Sciences Centre, Medical Research
             Council, Du Cane Rd, London W12 ONN, UK
FEATURES   Location/Qualifiers
     source   1..8372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q32-q33"
                     /clone="pCL4"
     mRNA   <1167..>1997
                     /gene="SOX21"
                     /product="DNA-binding protein SOX21"
     gene   <1167..>1997
                     /gene="SOX21"
     CDS   1167..1997
                     /gene="SOX21"
                     /note="Sry related HMG-box containing gene 21"
                     /codon_start=1
                     /product="DNA-binding protein SOX21"
                     /protein_id="AAC95381.1"
                     /db_xref="GI:4008103"
                     /translation="MSKPVDHVYKRPMNAFMVWSRAQRKMAQENPRKMHNSEISKRLGA
                     EWKLLTESEKRPFIDEAKRLRAMHMKEHPDYKYRPRRKPKTLLKKDKFAFPVPYGLGG
                     VADAEHPALKAGAGLHAGAGGGLVPESLLANPEKAAAAAAAARVFFPQSAAAAAAAA
                     AAAAAAQSPYISLLDIGSKMAEISSSSSSLPYASSLGYTPAGAFHGAAAAAAAAA
                     AGGHTHSHPSPGNPGVGVMIPCNCSAWPSPGIQPPLAYILLPGMGKRPQLDPYPAAYAAAL"

RESULT 3

BASE COUNT   1996 a   2143 c   2104 g   2124 t   5 others
ORIGIN

Query Match   100.0%; Score 500; DB 9; Length 8372;
Best Local Similarity   100.0%; Pred. No. 1.8e-144;
Matches 500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 AAGCTTGGTGCCATCTATTTTGGACTATGCCTTGCATACAGCTTTATGGGAACATTTGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTGGTGCCATCTATTTTGGACTATGCCTTGCATACAGCTTTATGGGAACATTTGTC 60

Qy     61 AGGCAAAAGTATATAATAAATGGCAAAACTCTACGCCCTTTTATTTTAAATTAGATTGGTGTGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGCAAAAGTATATAATAAATGGCAAAACTCTACGCCCTTTTATTTTAAATTAGATTGGTGTGAT 120

Qy    121 TTGATGCTGACGGGAGTGAGAGAGTAATGGCCCTTATCCTCGCGTGTGCTGAGGATGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGATGCTGACGGGAGTGAGAGAGTAATGGCCCTTATCCTCGCGTGTGCTGAGGATGG 180

Qy    181 CCTGGGTCTGCCACCCTCCTCGAGTAGCATTTTGCATGTGTAACAGGGGGTCTCCCCTCTGGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTGGGTCTGCCACCCTCCTCGAGTAGCATTTTGCATGTGTAACAGGGGGTCTCCCCTCTGGG 240

Qy    241 GCACAACAACAAAAGAGAGAAGTTGCTAAGGACAAGAAGCAGGTGCGGGAAAATGCATCTCCCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCACAACAACAAAAGAGAGAAGTTGCTAAGGACAAGAAGCAGGTGCGGGAAAATGCATCTCCCAT 300

Qy    301 TGGAACAGACGCCCTGGGCTTACTCCACATGGGCTGAGAGAGTGCTATGGCCAGTCCTCCCAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGAACAGACGCCCTGGGCTTACTCCACATGGGCTGAGAGAGTGCTATGGCCAGTCCTCCCAGA 360

Qy    361 GCTCTCGAGCTGCACTTGGGGGTGGGACAGTTGGGGGTTTCCCATGGGCGGCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTCTCGAGCTGCACTTGGGGGTGGGACAGTTGGGGGTTTCCCATGGGCGGCA 420

Qy    421 GAAAGCCAGCCAACTGCTGCCCAAAAATCACCCAGCCGATTGGGGGTTTCCCATGGGCGGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAAGCCAGCCAACTGCTGCCCAAAAATCACCCAGCCGATTGGGGGTTTCCCATGGGCGGCA 480

Qy    481 CCCTGCCCGGAGCCAAGAAG 500
          ||||||||||||||||||||
Db    481 CCCTGCCCGGAGCCAAGAAG 500
```

LOCUS   AL159970/c
DEFINITION   AL159970
             Human DNA sequence from clone RP11-140I19 on chromosome 13 Contains
             STSs and GSSS, complete sequence.
ACCESSION   AL159970
VERSION   AL159970.16   GI:11121011
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

LOCUS   AL159970   112929 bp   DNA   linear   PRI 26-FEB-2001

RESULT 2

```
BASE COUNT       867 a     1682 c     1570 g      758 t
ORIGIN
LOCUS       AB023180      4877 bp    mRNA    linear   PRI 16-JUN-1999
DEFINITION  Homo sapiens mRNA for KIAA0963 protein, complete cds.
ACCESSION   AB023180
VERSION     AB023180.1  GI:4589569
KEYWORDS    .
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
            SK plus clone:hj05936.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosawa,M.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE       Prediction of the coding sequences of unidentified human genes.
            XIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
JOURNAL     DNA Res. 6 (1), 63-70 (1999)
MEDLINE     99246063
REFERENCE   2 (bases 1 to 4877)
AUTHORS     Ohara,O., Nagase,T. and Kikuno,R.
TITLE       Direct Submission
JOURNAL     Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..4877
```

```
Qy          1 GCCCGAAACCCGGAAGTGAGCGGCGGCCAGCTCGGAGGCTCGGAGAAACAGGCGCGCGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCCGAAACCCGGAAGTGAGCGGCGGCCAGCTCGGAGGCTCGGAGAAACAGGCGCGCGGG 60

Qy         61 CTCCGCGCCCCGGCCGGGGCGCGGGACCCCGAGATCATGATGCTGCCGCCGCCACCGCCACCAC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCGCGCCCCGGCCGGGGCGCGGGACCCCGAGATCATGATGCTGCCGCCGCCACCGCCACCAC 120

Qy        121 GGAGCGAGAGCCCAGATAGACGCCCGGGGGTCCTGGAGTCCGGCGCCTGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGAGCGAGAGCCCAGATAGACGCCCGGGGGTCCTGGAGTCCGGCGCCTGC 180

Qy        181 TGCCCGGCGGAGGACCCCACCCCGCCGCCGCCGCCATGG 240
              |||||||||||||||||||||||||||||||||||||||
Db        181 TGCCCGGCGGAGGACCCCACCCCGCCGCCGCCGCCATGG 240

Qy        241 ACAGGGATTACCCGCAGCATGAAACCCCGCCAGCATGAACCCCGCCAGCCTCCTGTTACAGCCCCGCCGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACAGGGATTACCCGCAGCATGAAACCCCGCCAGCATGAACCCCGCCAGCCTCCTGTTACAGCCCCGCCGC 300

Qy        301 CCCTGCAGACGCCCATGCTCCACTGCCCCTACTGGAACACCTTCTCGCTGCCGCCCATACC 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCTGCAGACGCCCATGCTCCACTGCCCCTACTGGAACACCTTCTCGCTGCCGCCCATACC 360

Qy        361 CTGCCTTCTCCAGCGACAGCGCGCCCGTTCATGAGCTCCGCCTCCTTCTCTCGGCGAGCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCCTTCTCCAGCGACAGCGCGCCCGTTCATGAGCTCCGCCTCCTTCTCTCGGCGAGCCAGC 420
```

```
Qy    421  CCTGCCCAGACACCAGCTATGCCCCGTGGCCAGCCGCTTCCAGCTTGCCACCAAAGACCT  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CCTGCCCAGACACCAGCTATGCCCCGTGGCCAGCCGCTTCCAGCTTGCCACCAAAGACCT  480

Qy    481  GCGACTTTGCTCAGGACTCC  500
           ||||||||||||||||||||
Db    481  GCGACTTTGCTCAGGACTCC  500
```

RESULT 3

```
LOCUS       AC011553/c
DEFINITION  AC011553                   35848 bp    DNA     linear   HTG 23-APR-2001
            Homo sapiens chromosome 19 clone LLNLR-277D11, WORKING DRAFT
            SEQUENCE, 3 ordered pieces.
ACCESSION   AC011553
VERSION     AC011553.3  GI:7711540
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 35848)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 19
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 35848)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 6, 2000 this sequence version replaced gi:7650187.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 26813, R31543
            Center clone name: LLNLR-277D11
            -----
            Summary Statistics
            Consensus quality: 35124 bases at least Q40
            Consensus quality: 35398 bases at least Q30
            Consensus quality: 35534 bases at least Q20
            Estimated Insert size: 32650; agarose-fp estimation
            Estimated Insert size: 35748; sum-of-contigs estimation
            Quality coverage:  9.64 in Q20 bases; agarose-fp estimation
            Quality coverage: 10.55 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 3 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces
            is believed to be correct as given, however the sizes
            of the gaps between them are based on estimates that have
            provided by the submittor.
            This sequence will be replaced
            by the finished sequence as soon as it is available and
            the accession number will be preserved.
          *         1  18613: contig of 18613 bp in length
          *     18614  18713: gap of unknown length
          *     18714  19277: contig of 564 bp in length
          *     19278  19377: gap of unknown length
          *     19378  35848: contig of 16471 bp in length.
FEATURES             Location/Qualifiers
     source          1..35848
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone_lib="Lawrence Livermore human cosmid library LLNLR"
                     /clone="LLNLR-277D11"
BASE COUNT      7758 a  11001 c  10770 g   6119 t    200 others
ORIGIN
```

```
Query Match                                  44.4%;  Score 221.8;  DB 2;  Length 35848;
Best Local Similarity    91.4%;
Matches  235;  Conservative  0;  Mismatches 22;  Indels  0;  Gaps  0;

Qy     73  CCGGACCCGGCGCCCGAGATCATGATGCTGCCGCGCCACCGGCCGCCACGGAGGAGGAAGC  132
           | ||  |  | |         |||||||||||||||||||||||||||||||||||||||||
Db   5256  CCCGTGCTTCCCTGCAGATCATGATGCTGCCGCGCCACCGGCCGCCACGGAGGAGGAGC   5197

Qy    133  CCAGATAGACGCCCCCGGCGCCCCGGCGCCCCGGGTCCCTGGAGTCCCGGCGCCGGCCGAG  192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5196  CCAGATAGACGCCCCCGGCGCCCCGGCGCCCCGGGTCCCTGGAGTCCCGGCGCCGGCCGAG  5137

Qy    193  GACCCCACCCCCGCCTGCCGCCCCGATGCTTGCAGTGGGGCCCCCCCATGGACAGGGATTACC  252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5136  GACCCCACCCCCGCCTGCCGCCCCGATGCTTGCAGTGGGGCCCCCCCATGGACAGGGATTACC  5077

Qy    253  CGCAGAТAGAACCCCCGCCGCGGGGGGCAGCCTCCTGGTTACAGCCCCCTCCTGCCGCAGAGCG  312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5076  CGCAGAТAGAACCCCGCCGCGGGGGGCAGCCTCCTGGTTACAGCCCCGCCCTGCCGCAGGTGA  5017

Qy    313  CCATGCTGCCACTGCCCC  329
           | ||  |  | ||||
Db   5016  GCCCCACCCCGCGCCCC   5000
```

RESULT 4

```
LOCUS       AC093067/c
DEFINITION  AC093067                   6153 bp    DNA     linear   PRI 24-OCT-2001
            Homo sapiens chromosome 19 clone LLNLR-240D7, complete sequence.
ACCESSION   AC093067
VERSION     AC093067.2  GI:16356878
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6153)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (09-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   2  (bases 1 to 6153)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 6153)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-2001) DOE Joint Genome Institute and Stanford
            Human Genome Center and Los Alamos National Laboratory.
COMMENT     On Oct 24, 2001 this sequence version replaced gi:15144408.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 98.4% of Sequence;
            Estimated Total Number of Errors is 0.
            NOTE: This sequence is not the entire sequence of the clone. It
            is sequence generated to span the gap between AC005390 and
            AC011553. The overlap with AC005390 is 1172bp and the overlap with
            AC011553 is 1062bp. The sequence was finished by the Stanford
            Human Genome Center and Los Alamos National Laboratory.
FEATURES             Location/Qualifiers
     source          1..6153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="LLNLR-240D7"
BASE COUNT      1346 a   2032 c   1887 g    888 t
ORIGIN

Query Match                 22.9%;  Score 114.4;  DB 9;  Length 6153;
Best Local Similarity   99.1%;  Pred. No. 8.7e-09;
```

```
LOCUS       AX201593                1872 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 14 from Patent WO0153349.
ACCESSION   AX201593
VERSION     AX201593.1  GI:15391442
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1872)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.Y.
  TITLE     Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 14 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..1872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      300 a    673 c    523 g    376 t
ORIGIN
```

Query Match   100.0%; Score 500; DB 6; Length 1872;
Best Local Similarity 100.0%; Pred. No. 2.3e-95;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60

Qy      61 CCTCCCCACCCTCTCTCCCAAGGCCCCTCTCCCTCGGTCTCCCTTCTTCCTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCTCCCCACCCTCTCTCCCAAGGCCCCTCTCCCTCGGTCTCCCTTCTTCCTC 120

Qy     121 CACCTCCCTCTCTGCAGAACTTCTTCCTTTACCCCCACCACTGCCCCCCTTTCC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CACCTCCCTCTCTGCAGAACTTCTTCCTTTACCCCCACCACTGCCCCCCTTTCC 180

Qy     181 TTTTTCTGACCTCCTTTTGAGGGGCTCACCGCTCGCCCAGACCATAGGAGAGAGAGATGTGGGAGG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTTTCTGACCTCCTTTTGAGGGGCTCACCGCTCGCCCAGACCATAGGAGAGAGAGATGTGGGAGG 240

Qy     241 CTCCAGTTCCTGGGCTTGCTGTGTTTCTGCACGCGCTTTGGGTGTGGCTTCCCAGTGAAGCCTCTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCCAGTTCCTGGGCTTGCTGTGTTTCTGCACGCGCTTTGGGTGTGGCTTCCCAGTGAAGCCTCTCC 300

Qy     301 ACCCAGGGGGCTGAGGTCCCGGTGGTGTGGGGCCCAGGAGGGGGGCTCCTGCCCAGCTCCCCT 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACCCAGGGGGCTGAGGTCCCGGTGGTGTGGGGCCCAGGAGGGGGGCTCCTGCCCAGCTCCCCT 360

Qy     361 GCAGCCCCACCAATCCCCTTCCAGGATCTCGAAGACGCAGGGGTCACTTGGC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCAGCCCCACCAATCCCCTTCCAGGATCTCGAAGACGCAGGGGTCACTTGGC 420

Qy     421 AGCATCAGCCAGACGAGTGGCCCGCCCCGCCGCCCCCGGGCCCATCCCCTGCCCCCCGGGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGCATCACCAGACAGTGGCCCGCCCCGCCGCCCCCGGGCCCATCCCCTGCCCCCCGGGCC 480

Qy     481 CTCACCCGGCGGGGCGCCCTCC 500
           |||||||||||||||||||||||
Db     481 CTCACCCGGCGGGGCGCCCTCC 500
```

```
RESULT 9
HSLAG3
LOCUS       HSLAG3                   1872 bp    mRNA    linear   PRI 12-SEP-2001
DEFINITION  Human LAG-3 mRNA for CD4-related protein involved in lymphocyte
            activation.
ACCESSION   X51985
VERSION     X51985.3  GI:15617340
KEYWORDS    cell surface glycoprotein; immune response; immunoglobulin
```

superfamily; transmembrane protein.

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1872)
  AUTHORS   Triebel,F., Jitsukawa,S., Baixeras,E., Roman-Roman,S., Genevee,C.,
            Viegas-Pequignot,E. and Hercend,T.
  TITLE     LAG-3, a novel lymphocyte activation gene closely related to CD4
  JOURNAL   The Journal of experimental medicine. 171 (5), 1393-1405 (1990)
   MEDLINE   90237736
REFERENCE   2  (bases 1 to 1872)
  AUTHORS   Triebel,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1990) Triebel F., Laboratoire d'Immunologie
            Cellulaire U333, Institut Gustave Roussy rue Camille Desmoulins,
            94805 Villejuif, France
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 1872)
  AUTHORS   Triebel,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1996) Triebel F., Laboratoire d'Immunologie
            Cellulaire U333, Institut Gustave Roussy rue Camille Desmoulins,
            94805 Villejuif, France
  REMARK    Revised by [4]
REFERENCE   4  (bases 1 to 1872)
  AUTHORS   Triebel,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) Triebel F., Laboratoire d'Immunologie
            Cellulaire U333, Institut Gustave Roussy rue Camille Desmoulins,
            94805 Villejuif, France
COMMENT     On Sep 13, 2001 this sequence version replaced gi:11558021.
            Data kindly reviewed (08-OCT-1990) by Triebel F.
FEATURES             Location/Qualifiers
     source          1..1872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MB-F(5)"
     sig_peptide     231..296
                     /note="signal peptide"
     CDS             231..1808
                     /codon_start=1
                     /product="LAG-3 protein precursor"
                     /protein_id="CAA36243.3"
                     /db_xref="GI:15617341"
                     /db_xref="SWISS-PROT:P18627"
                     /translation="MWEAQFLGLLFLQPLWVAPVKPLQPGAEVPVVWAQEGAPAQLPC
                     SPTIPLQDLSLLRRAGVTWQHQPDSGPPAAAPGHPLAPGPHPAAPSSWGPRPRRTYVL
                     SVGPGGLRSGRLPLQPRVQLDERGRQRDPSLWLRPARRADAGEYRAAVHLRDRALSC
                     RLRLRLGQASMTASPPGSLRASDWVILNCSFSRPDRPASVHWFRNRGQGRVPVRESPH
                     HHLAESFLFLPQVSPMDSGPWGCILTYRDGFNVSIMYNLTVLGLEPPTPLTVYAGAGS
                     RVGLPCRLPAGVGTRSFLTAKWTPPGGGPDLLVTGDNGDFTLRLEDVSQAQAGTYCH
                     IHLQQQLNATVTLAIITVVTPKSFGSPGSILGKLLCEVTPVSGQERFVWSSLDTPSQRS
                     FSGPWLEAQEAQLLSQPWQCQLYQGERLLGAAVYFTELSSPGAQRSGRAPGALPAGHL
                     LFLTIGVLSLILLLVTGAFGFHLWRRQWRPRRFSALEQGIHPPQAQSKIELEQEPEP
                     EPEPEPEPEPEPEQL"
     mat_peptide     297..1805
                     /product="LAG-3 protein"
BASE COUNT      300 a    673 c    523 g    376 t
ORIGIN
```

Query Match   100.0%; Score 500; DB 9; Length 1872;
Best Local Similarity 100.0%; Pred. No. 2.3e-95;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60

Qy      61 CCTCCCCACCCTCTCTCCCAAGGCCCCTCTCCCTCGGTCTCCCTTCTTCCTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCTCCCCACCCTCTCTCCCAAGGCCCCTCTCCCTCGGTCTCCCTTCTTCCTC 120
```

RESULT 10

```
LOCUS       A81356          2279 bp    DNA     linear   PAT 21-JAN-2000
DEFINITION  Sequence 1 from Patent EP0900841.
ACCESSION   A81356
VERSION     A81356.1  GI:6731677
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2279)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE     LAG-3 splice variants
  JOURNAL   Patent: EP 0900841-A 1 10-MAR-1999;
            APPLIED RESEARCH SYSTEMS (AN); INST NAT SANTE RECH MED (FR)
FEATURES             Location/Qualifiers
     source          1..2279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      360 a    832 c    589 g    498 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 2279;
  Best Local Similarity 100.0%;  Pred. No. 2.2e-95;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACCACTGCCCCCTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACCACTGCCCCCTTTCC 180

Qy        181 TTTTCTCGACCTCCTTTTGGAGGGCTCAGCGGTCAGCGCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTCTCGACCTCCTTTTGGAGGGCTCAGCGGTCAGCGCTCA 240

Qy        241 CTCAGTTCCTGGGGCTTGCTGTTCTGCAGCGCGTGAAAGCCTCTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAGTTCCTGGGGCTTGCTGTTCTGCAGCGCGTGAAAGCCTCTCC 300

Qy        301 AGCCAGGGGCTGAGGTCCCGGTGGTGTGGGGCCCAGGAGGGGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCAGGGGCTGAGGTCCCGGTGGTGTGGGGCCCAGGAGGGGGG 360

Qy        361 GCAGCCCCACACAATCCCCTCCAGGATCTCAGCGCTTCTGCGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCAGCCCCACACAATCCCCTCCAGGATCTCAGCGCTTCTGCGGA 420

Qy        421 AGCATCAGCCAGAGACAGTGGCCCGCCGGCCCATCCCCTGGCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCATCAGCCAGAGACAGTGGCCCGCCGGCCCATCCCCTGGCCC 480

Qy        481 CTCACCCGCGGGCGCGCCCTCC 500
              ||||||||||||||||||||||
Db        481 CTCACCCGCGGGCGCGCCCTCC 500
```

RESULT 11

```
LOCUS       AX001576        2279 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 1 from Patent WO9858059.
ACCESSION   AX001576
VERSION     AX001576.1  GI:7241708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2279)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE     LAG-3 SPLICE VARIANTS
  JOURNAL   Patent: WO 9858059-A 1 23-DEC-1998;
            INST NAT SANTE RECH MED (FR); ROUSSY INST GUSTAVE (FR)
FEATURES             Location/Qualifiers
     source          1..2279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      360 a    832 c    589 g    498 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 2279;
  Best Local Similarity 100.0%;  Pred. No. 2.2e-95;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60

Qy         61 CCTCCCCACCCCTCTCCAAAGGCCCTCCTCCTTGGTCTCCCTGGTCGTCTTCCTCTGGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCTCCCCACCCCTCTCCAAAGGCCCTCCTCCTTGGTCTCCCTGGTCGTCTTCCTCTGGAT 120

Qy        121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACCACTGCCCCCTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACCACTGCCCCCTTTCC 180

Qy        181 TTTTCTCGACCTCCTTTTGGAGGGCTCAGCGGTCAGCGCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTCTCGACCTCCTTTTGGAGGGCTCAGCGGTCAGCGCTCA 240

Qy        241 CTCAGTTCCTGGGGCTTGCTGTTCTGCAGCGCGTGAAAGCCTCTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCAGTTCCTGGGGCTTGCTGTTCTGCAGCGCGTGAAAGCCTCTCC 300

Qy        301 AGCCAGGGGCTGAGGTCCCGGTGGTGTGGGGCCCAGGAGGGGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCCAGGGGCTGAGGTCCCGGTGGTGTGGGGCCCAGGAGGGGGG 360

Qy        361 GCAGCCCCACACAATCCCCTCCAGGATCTCAGCGCTTCTGCGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCAGCCCCACACAATCCCCTCCAGGATCTCAGCGCTTCTGCGGA 420

Qy        481 CTCACCCGCGGGCGCGCCCTCC 500
              ||||||||||||||||||||||
Db        481 CTCACCCGCGGGCGCGCCCTCC 500
```

```
BASE COUNT       166 a      261 c      246 g      72 t      3 others
ORIGIN

Query Match
Best Local Similarity  100.0%;  Score 500;  DB 6;  Length 748;
Matches 500;  Conservative  100.0%;  Pred. No. 1.8e-36;
Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGTCTACGGACCATTGCTGAGGAGGACGGCTGCCCAGGGACGCGTTGCCCAAGTCGTGCTTAATGAGGAGGACGCCGACCAGG  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGACCATTGCTGAGGACGCTGCCCAGGGACGCGTTGCCCAAGTCGTGCTTAATGAGGAGGACGCCGACCAGG       60

Qy     61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCAACGAGGAAGCCGCCCAAGGGGCATCG                            120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCAACGAGGAAGCCGCCCAAGGGGCATCG                            120

Qy    121 CCGAGGACGCCATCCAGGGCATCGGCCAACGAGGAGGTTGCCCAAGTCGCCAATGGGG                                180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGGACGCCATCCAGGGCATCGGCCAACGAGGAGGTTGCCCAAGTCGCCAATGGGG                                180

Qy    181 TCGCCCGCACAGGGCATCGGCCAATGAGGAGGACGCCGTGGGACGCCG                                          240
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGCCCGCACAGGGCATCGGCCAATGAGGAGGACGCCGTGGGACGCCG                                          240

Qy    241 TCCACGGCGTTCGCCAACGGGGGACGCCGTTCCTCAGCTTCGCCAACGGGGGACGCCCAGG                             300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCACGGCGTTCGCCAACGGGGGACGCCGTTCCTCAGCTTCGCCAACGGGGGACGCCCAGG                             300

Qy    301 GCATCGCCAACGAGGGGGACGCCACCAAGGGCATCGCCAAGGGCAATGGGG                                       360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCATCGCCAACGAGGGGGACGCCACCAAGGGCATCGCCAAGGGCAATGGGG                                       360

Qy    361 CTAACGAGGAGACGCCGTCCAGGGCATCGGCCATTCCACGGCGCGCCCAGG                                       420
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTAACGAGGAGACGCCGTCCAGGGCATCGGCCATTCCACGGCGCGCCCAGG                                       420

Qy    421 AGGACGCCGCCCAGGGAATCGCCGAGGATGTCGCACAGGGCATCGCCAACACAAG                          500
          |||||||||||||||||||||||||||||||||||||||||
Db    421 AGGACGCCGCCCAGGGAATCGCCGAGGATGTCGCACAGGGCATCGCCAACACAAG                          480

Qy    481 CCCAGGGCCATCGCCCAACACAAG                                                          500
          |||||||||||||||||||||||
Db    481 CCCAGGGCCATCGCCCAACACAAG                                                          500
```

```
RESULT 2
LOCUS        AX201594            1201 bp    DNA        linear    PAT 30-AUG-2001
DEFINITION   Sequence 15 from Patent WO0153349.
ACCESSION    AX201594
VERSION      AX201594.1  GI:15391443
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1201)
AUTHORS      Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
             Chen,Y.T.
TITLE        Small cell lung cancer associated antigens and uses therefor
JOURNAL      Patent: WO 0153349-A 15 26-JUL-2001;
             LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
             SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
             INC. (US)

FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
RESULT 3
LOCUS        HSM801408           2448 bp    mRNA        linear    PRI 18-FEB-2000
DEFINITION   Homo sapiens mRNA; cDNA DKFZp434C196 (from clone DKFZp434C196);
             partial cds.
ACCESSION    AL133561
VERSION      AL133561.1  GI:6599133
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2448)
AUTHORS      Koehler,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE        Direct Submission
JOURNAL      Submitted (15-DEC-1999) MIPS, Am Klopferspitz 18a, D-82152
             Martinsried, GERMANY
COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Resarch Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
             Clone from DKFZ; Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
             information about the clone and the sequencing project is available
             at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES             Location/Qualifiers
     source          1..2448
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434C196"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
```

```
Qy      1 GAGTCTACGGACCATTGCTGAGGAGGACGGCTGCCCAGGGACGCGTTGCCCAAGTCGTGCTTAATGAGGAGGACGCCGACCAGG  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGACCATTGCTGAGGACGCTGCCCAGGGACGCGTTGCCCAAGTCGTGCTTAATGAGGAGGACGCCGACCAGG       60

Qy     61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCAACGAGGAAGCCGCCCAAGGGGCATCG                            120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCAACGAGGAAGCCGCCCAAGGGGCATCG                            120

Qy    121 CCGAGGACGCCATCCAGGGCATCGGCCAACGAGGAGGTTGCCCAAGTCGCCAATGGGG                                180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGGACGCCATCCAGGGCATCGGCCAACGAGGAGGTTGCCCAAGTCGCCAATGGGG                                180

Qy    181 TCGCCCGCACAGGGCATCGGCCAATGAGGAGGACGCCGTGGGACGCCG                                          240
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGCCCGCACAGGGCATCGGCCAATGAGGAGGACGCCGTGGGACGCCG                                          240

Qy    241 TCCACGGCGTTCGCCAACGGGGGACGCCGTTCCTCAGCTTCGCCAACGGGGGACGCCCAGG                             300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCACGGCGTTCGCCAACGGGGGACGCCGTTCCTCAGCTTCGCCAACGGGGGACGCCCAGG                             300

Qy    301 GCATCGCCAACGAGGGGGACGCCACCAAGGGCATCGCCAAGGGCAATGGGG                                       360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCATCGCCAACGAGGGGGACGCCACCAAGGGCATCGCCAAGGGCAATGGGG                                       360

Qy    361 CTAACGAGGAGACGCCGTCCAGGGCATCGGCCATTCCACGGCGCGCCCAGG                                       420
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTAACGAGGAGACGCCGTCCAGGGCATCGGCCATTCCACGGCGCGCCCAGG                                       420

Qy    421 AGGACGCCGCCCAGGGAATCGCCGAGGATGTCGCACAGGGCATCGCCAACACAAG                            480
          |||||||||||||||||||||||||||||||||||||||||
Db    421 AGGACGCCGCCCAGGGAATCGCCGAGGATGTCGCACAGGGCATCGCCAACACAAG                            480

Qy    481 CCCAGGGCCATCGCCCAACACAAG                                                            500
          |||||||||||||||||||||||
Db    481 CCCAGGGCCATCGCCCAACACAAG                                                            500
```

```
BASE COUNT       255 a      377 c      349 g      178 t      42 others
ORIGIN

Query Match
Best Local Similarity  100.0%;  Score 500;  DB 6;  Length 1201;
Matches 500;  Conservative  100.0%;  Pred. No. 1.6e-36;
Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
RESULT 4
LOCUS           AL356585            184590 bp    DNA     linear   HTG 20-JAN-2001
DEFINITION      Homo sapiens chromosome 13 clone RP11-341D18, *** SEQUENCING IN
                PROGRESS ***, 13 unordered pieces.
ACCESSION       AL356585
```

VERSION   AL356585.3 GI:9800966
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 184590)
  AUTHORS   Burton,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Aug 14, 2000 this sequence version replaced gi:9213941.
COMMENT   -------------- Summary Statistics
            Center project name: bA341D18
            -------------- Project Information
            Contact: humquery@sanger.ac.uk
            Web site: http://www.sanger.ac.uk
            Center code: SC
            Center: Sanger Centre
            -------------- Genome Center
            as soon as it is available and the accession number will
            This record will be updated with the finished sequence
            runs of N, but the exact sizes of the gaps are unknown.
            arbitrary. Gaps between the contigs are represented as
            is not known and their order in this sequence record is
            consists of 13 contigs. The true order of the pieces
          * NOTE: This is a 'working draft' sequence. It currently
          * be preserved.
coverage: 4.20x in Q20 bases; agarose-fp
Quality coverage: 4.32x in Q20 bases; agarose-fp Quality
Insert size: 192611; 2.2% error; agarose-fp
Insert size: 183390; sum-of-contigs
Consensus quality: 182219 bases at least Q20
Consensus quality: 181067 bases at least Q30
Consensus quality: 179028 bases at least Q40
Chemistry: Dye-terminator Big Dye; L08752; 100% of reads
Sequencing vector: plasmid; pUC18; L08752; 100% of reads
Assembly program: XGAP4; version 4.5

FEATURES   Location/Qualifiers
     source   1..184590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-341D18"
                     /clone_lib="RPCI-11.2"
     misc_feature   1..36485
```
            1      36485: contig of 36485 bp in length
          36486      36585: gap of           100 bp
          36586      61226: contig of 24641 bp in length
          61227      61326: gap of           100 bp
          61327      86064: contig of 24738 bp in length
          86065      86164: gap of           100 bp
          86165      92534: contig of 6370 bp in length
          92535      92634: gap of           100 bp
          92635      97289: contig of 4655 bp in length
          97290      97389: gap of           100 bp
          97390     100517: contig of 3128 bp in length
         100518     100617: gap of           100 bp
         100618     116404: contig of 15787 bp in length
         116405     116504: gap of           100 bp
         116505     124935: contig of 8431 bp in length
         124936     125035: gap of           100 bp
         125036     127757: contig of 2722 bp in length
         127758     127857: gap of           100 bp
         127858     140183: contig of 12326 bp in length
         140184     140283: gap of           100 bp
         140284     143272: contig of 2989 bp in length
         143273     143372: gap of           100 bp
         143373     156929: contig of 13557 bp in length
         156930     157029: gap of           100 bp
         157030     184590: contig of 27561 bp in length

```
gene            1..1744
                /gene="DKFZp434C196"
CDS             <1..1744
                /gene="DKFZp434C196"
                /note="similar to mucins"
                /codon_start=2
                /product="hypothetical protein"
                /protein_id="CAB63715.1"
                /db_xref="GI:6599134"
                /translation="SPSRASLTRTPPRASLMRRPSTASLTRTPSRASPTRMPESRASLK
                MTFPRASLTKMESTALLRTLPRASLMKRTPPTRASLMRRPSTASLTRTPSRASPTRKPPRASRTPSRA
                SPTRRLPRASPMGSPHRASPMRTPPRASPTGTPSTASPTGTPSSASPTGTPPRASPTG
                TPPRAWATRPSPSTASLTRTPSRASLTRWPPRASPTGTPTPPESPRMSHRASPPRTPPRA
                SPTRRPPRASPTPRTPPRESLRTISHRASPTRMPEPRASPTRRPPRASPTGSSPPRASPMTP
                PRASPTPPRASPTTPSRASLTRTPSWASPTRPSRASLMKMESTVSITRTPPRASP
                TGTPSRASPTGTPSRASLTGPSRASLIGTPSRASLIGTPSRASLIGTP
                PRASLTGTSSTASLTRTPSRASLIGTPSRASLIGTPSRASLTRTPPRAS
                LTRTPRASLTRTPPRASLTRTPSMVLLKRSPSRASLTRTPSRASLTMTPSRASLTRT
                PSTASLIRTGTPPPIASLTRTPPTASLTRSPPTASLTRTPTASLTRMPSTASLTRKSNVN
                QCPASTPSSEVIS"
```

polyA_signal   2397..2402
polyA_site   2418

BASE COUNT   643 a   727 c   648 g   430 t
ORIGIN

Query Match   99.7%; Score 498.4; DB 9; Length 2448;
Best Local Similarity   99.8%; Pred. No. 1.7e-36;
Matches 499; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy          1 GAGHCTACGGCATTGCTGAGGAGGGCTGCCCAGGGCCATCCGCTAATGAGGACGCCCAAGG 60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        167 GAGTCTACGGCATTGCTGAGGACGGCTGCCCAGGGCCATCCGCTAATGAGGACGCCCAAGG 226

Qy         61 GCATCGCTAATGAGGACACCACCCCAGTGCATCGCCAACGAGGAAGCCGCCCCAGGGCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GCATCGCTAATGAGGACACCACCCCAGGGCATCGCCAACGAGGAAGCCGCCCCAGGGCATCG 286

Qy        121 CCGAGGACGGCCATCCAGGGCATCGCCAACGAGGAGGAGGTTGCCCAGGGCATCGCCAATGGGG 180
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        287 CCGAGGACGGCCATCCAGGGCATCGCCAACGAGGAGGTTGCCCAGGGCATCGCCAATGGGG 346

Qy        181 TCGCCGCACAGGGCATCGCCAATGAGGACCCCCACCCCAGGGCATCGCCAACTGGGACGCG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 TCGCCCCACAGGGCATCGCCAATGAGGACGCCCACCCCAGGGCATCGCCAACTGGGACGCG 406

Qy        241 TCCACGGGCTTCGCCAACGGGGACGCGCGTCCTCAGCTTCGCCAACGGGGACGCGCGCCAGG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TCCACGGGCTTCGCCAACGGGGACGCCGTCCTCAGCTTCGCCAACGGGGACGCGCGCCAGG 466

Qy        301 GCATCGCCAACGGGGACGGCCACCAAGGCCATGGGCACGAGGTTCACCATCCACGGCATCG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 GCATCGCCAACGGGGACGCCCACCAGGGCATGGGCACGAGGTCACCATCCACGGCATCG 526

Qy        361 CTAACGAGGACGGCCGCGTCCAGGGCCGCGTTACGGAGTCGCCAACGGGCCATCGCCAACG 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCGCGTGGCCCATCGCCAACG 586

Qy        421 AGGACGCGCGCCCAGGGAATCGCCGAGGATGTCGCACAGGGCACGCCAACGAGGACGCG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        587 AGGACGCCGCGCCCAGGGAATCGGCGGGATGTCGCACAGGGCATCGCCAACGAGGACGCG 646

Qy        481 CCCAGGGGCCATCGCCAACAAG 500
              ||||||||||||||||||||||
Db        647 CCCAGGGCCATCGCCAACAAG 666
```

```
\SE COUNT     166 a     261 c     246 g     72 t     3 others
\RIGIN

Query Match
Best Local Similarity   100.0%;     Score 500;  DB 6;   Length 748;
Matches 500; Conservative 0; Mismatches 0;  Indels 0;  Gaps 0;

Qy     1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGACGCCGACCAGG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGACGCCGACCAGG 60

Qy    61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCCAGGGCATCGCCCAGGGCATCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCCAGGGCATCGCCCAGGGCATCG 120
```

```
Qy     1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGACGCCGACCAGG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGACGCCGACCAGG 60

Qy    61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCCAGGGCATCGCCCAGGGCATCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GCATCGCTAATGAGGACACCACCCAGTGCATCGCCCAGGGCATCGCCCAGGGCATCG 120

Qy   121 CCGAGGACGCCATCCAGGGCATCGCCCAACCAGTTGTGCCCAGGGCATGGGG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCGAGGACGCCATCCAGGGCATCGCCCAACCAGTTGTGCCCAGGGCATGGGG 180

Qy   181 TCGCCCGACAGGGCATCGCCAATGAGGAGGACGCCGGCGCGCGCCGCCCCAG 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCGCCCGACAGGGCATCGCCAATGAGGAGGACGCCGGCGCGCGCCGCCCCAG 240

Qy   241 TCCACGGCTTCGCCAACGGGGACGCCGTCTTCAGCTTCGCCAACGGGCATCG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TCCACGGCTTCGCCAACGGGGACGCCGTCTTCAGCTTCGCCAACGGGCATCG 300

Qy   301 GCATCGCCAACGGGCATGGGCCATCGGGGCAACGAGGTCACCATCCACGGCATCG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCATCGCCAACGGGCATGGGCCATCGGGGCAACGAGGTCACCATCCACGGCATCG 360

Qy   361 CTAACGAGGAGGACGCCGTCCAGGGCCATCCGCCGGGCATCGCCCACG 420
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTAACGAGGAGGACGCCGTCCAGGGCCATCCGCCGGGCATCGCCCACG 420

Qy   421 AGGACGCCGCCCCAGGGAATCGCCGAGGATGTCGCACAGGGACGCCG 480
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AGGACGCCGCCCCAGGGAATCGCCGAGGATGTCGCACAGGGACGCCG 480

Qy   481 CCCAGGGCCATCGCCAACAAG 500
         ||||||||||||||||||||
Db   481 CCCAGGGCCATCGCCAACAAG 500
```

```
gene            /dev_stage="adult"
                1..1744
                /gene="DKFZp434C196"
CDS             <1..1744
                /gene="DKFZp434C196"
                /note="similar to mucins"
                /codon_start=2
                /product="hypothetical protein"
                /protein_id="CAB63715.1"
                /db_xref="GI:6599134"
                /translation="SPSRASLTRPPRASLMRRPSTASLTRTPSRASPTRMPSRASLK
                SPTRLRPASPMGSSPHRASPMRTPPRASPTGTPSTPSSASPTGPTPRASPTG
                MFFRASLTKMESTALLRFLPRASLMRTPPRASLMRTPPRASPTRKPPRASPTRPPSRA
                TPPRAWATRSPTRASLTRWPPRASPTRPRESPRMSHRASPTRTPPRA
                SPTRRPPRASPTRTPPRESLRTSHRASPTRMPPRASPTRRPRASPTGSPPRASPMTP
                PRASPRPPRASPTTTPSRASLTRTPSWASPTTTPSRASLMKMESTVSITRTPPRASP
                TGTPSRASPTGTPSRASLTRTPSRASLIGTPSRASLIGTPSRASLTGTP
                PRASLTGTSTASLTRTPSRASLTRTQSSSSLTRTPSMASLTRTPPRASLTRTPPRAS
                LIRTPPPRASLTRTPPRASLTRTPSMVSLKRSPSRASLTRTPSRASLTMPSRASLIRT
                PSTASLTRTSPPPRASLTRTPPTASLTRTPSPPTASLTRFPSTASLTRMPSTASLTRKSNVN
                QQCPASTPSSEVIS"
polyA_signal    2397..2402
polyA_site      2418
BASE COUNT      643 a    727 c    648 g    430 t
ORIGIN
```

```
Query Match     99.7%;  Score 498.4;  DB 9;  Length 2448;
Best Local Similarity  99.8%;  Pred. No. 1.7e-36;
Matches 499;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 GAGYCTACGGCCATTGCTGAGGAGCGTGCCCAGGACGCATCGCTAATGAGGACGCCGACCAGG 60
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  167 GAGTCTACGGCATTGCTGAGGACGCTGCCCCAGGCATCGCTAATGAGGACGCCGACCAGG 226

Qy   61 GCATCGCTAATGAGGACACCACCCACTGCATCGCCAACGAGAAGCGCGCCCCAGGCATCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  227 GCATCGCTAATGAGGACACCACCCACTGCATCGCCAACGAGAAGCGCGCCCCAGGCATCG 286

Qy  121 CCGAGGACGCCATCCAGGACGCATCGCCAACGAGGAGGGTTGCCCAGGCATCGCCAATGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  287 CCGAGGACGCCATCCAGGACGCATCGCCAACGAGGAGGGTTGCCCAGGCATCGCCAATGGGG 346

Qy  181 TCGCCCCACAGGGCATCGCCAATGAGGAACGCCACCCAGGCATCGCCAACTGGGACGCGCG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  347 TCGCCGCCACAGGGCATCGCCAATGAGGACGCGCCACCCAGGCATCGCCAACTGGGACGCGCG 406

Qy  241 TCCACGGCTTCGCCAACGAGGACGACGCCGTCCTCCAGCTTCGCCAACGAGGGGACGGCGCCCCAGG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  407 TCCACGGCTTCGCCAACGAGGACGCCGTCCTCAGCTTCGCCAACGAGGGGACGCGCCCCAGG 466

Qy  301 GCATCGCCAACGGGGACGGCGACGCCCACCAAGGGCATCTGGGCAACGAGGGTCACCCATCCACGGCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  467 GCATCGCCAACGGGGACCCACCACAAGGGCATCTGGGCAACGAGGGTCACCATCCACGGCATCG 526

Qy  361 CTAACGAGGAGACGCGCCGTCCAGGGCGGATCGCTAACGAGGTGCGGCCGCCCAGGGCATCGCCAACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  527 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGCGGCCGCCCAGGGCATCGCCAACG 586

Qy  421 AGGACGCCGCCCACAGGGAATCGCCCGAGATGTCGCACGGGGCATCGCCAACGAGGACGCCG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  587 AGGACGCCGCCCACAGGGAATCGCCGAGGATGTCGCACAGGGCATCGCCAACGAGGACGCGCG 646

Qy  481 CCCAGGGGCGCATCGCCAACAAG 500
          |||||||||||||||||||||||
Db  647 CCCAGGGGCATCGCCAACAAG 666
```

```
RESULT 4
LOCUS       AL356585
DEFINITION  Homo sapiens chromosome 13 clone RP11-341D18, *** SEQUENCING IN
            PROGRESS ***, 13 unordered pieces.
ACCESSION   AL356585
```

```
VERSION     AL356585.3  GI:9800966
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184590)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequests@sanger.ac.uk
            On Aug 14, 2000 this sequence version replaced gi:9213941.
COMMENT     -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA341D18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing program: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Sequencing vector: plasmid; L08752; 100% of reads
            Consensus quality: 181067 bases at least Q40
            Consensus quality: 179028 bases at least Q40
            Consensus quality: 182219 bases at least Q30
            Insert size: 183390; sum-of-contigs
            Insert size: 183611; 2.2% error; agarose-fp
            Insert size: 192611; 2.2% error; agarose-fp
            Quality coverage: 4.32x in Q20 bases; sum-of-contigs Quality
            coverage: 4.20x in Q20 bases; agarose-fp
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *    1           36485: contig of 36485 bp in length
            *    36486       36585: gap of         100 bp
            *    36586       61226: contig of 24641 bp in length
            *    61227       61326: gap of         100 bp
            *    61327       86064: contig of 24738 bp in length
            *    86065       86164: gap of         100 bp
            *    86165       92534: contig of 6370 bp in length
            *    92535       92635: gap of         100 bp
            *    92635       97289: contig of 4655 bp in length
            *    97290       97389: gap of         100 bp
            *    97390      100517: contig of 3128 bp in length
            *   100518      100617: gap of         100 bp
            *   100618      116404: contig of 15787 bp in length
            *   116405      116504: gap of         100 bp
            *   116505      124935: contig of 8431 bp in length
            *   124936      125035: gap of         100 bp
            *   125036      127757: contig of 2722 bp in length
            *   127758      127857: gap of         100 bp
            *   127858      140183: contig of 12326 bp in length
            *   140184      140283: gap of         100 bp
            *   140284      143272: contig of 2989 bp in length
            *   143273      143372: gap of         100 bp
            *   143373      156929: contig of 13557 bp in length
            *   156930      157029: gap of         100 bp
            *   157030      184590: contig of 27561 bp in length.
FEATURES             Location/Qualifiers
     source          1..184590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-341D18"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..36485
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: October 10, 2002, 12:39:37

Title: US-09-489-101A-3_COPY_1_500
Perfect score: 500
Sequence: 1 cacagcgcccgcatgtacaa.........atggacagttacgcgcacat 500

Scoring table: IDENTITY_NDC
                 Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: gb_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_hgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length DB	ID	Description

1	500	100.0	1085	6	AX201582	AX201582 Sequence
2	500	100.0	1085	9	HSSOX2G	231560 H.sapiens s
3	500	100.0	1109	9	HUMHMGBOX	L07335 Homo sapien
4	500	100.0	1181	9	BC013923	BC013923 Homo sapi
5	496.8	99.4	832	9	HSA327010	AJ327010 Homo sapi
6	449	89.8	3054	4	OASOX2GEN	X96997 O.aries SOX
7	447.8	89.6	749	9	HSA324649	AJ324649 Homo sapi
8	439.2	87.8	203344	2	AL606746	AL606746 Mus muscu
9	436	87.2	2418	10	MMSOX2	X94127 M.musculus
10	432.8	86.6	2283	10	AX305428	AX305428 Sequence
11	432.8	86.6	2283	10	MMU31967	U31967 Mus musculu
12	359.2	71.8	1186	5	CHKSOX2	D50603 Chicken mRN
13	359.2	71.8	1355	5	GGU12532	U12532 Gallus gall
14	350	70.0	395	6	AX321190	AX321190 Sequence
15	288.8	57.8	977	5	AF022928	AF022928 Xenopus 1
16	287.2	57.4	1190	5	AF005476	AF005476 Xenopus 1
17	279	55.8	1542	5	AX001335	AX001335 Sequence
18	279	55.8	4091	6	AX201583	AX201583 Sequence
19	279	55.8	4091	9	HSSOX1	Y13436 Homo sapien
20	279	55.8	147665	9	AL138691	AL138691 Human DNA
c 21	266.2	53.2	2312	6	AB011802	AB011802 Gallus ga
22	266.2	53.2	2312	10	AX001333	AX001333 Sequence
23	262.8	52.6	1150	10	MMSOX3	X94125 M.musculus
24	262.8	52.6	18632	10	AF434675	AF434675 Mus muscu
25	261.4	52.3	2376	10	AX001334	AX001334 Sequence
26	261.4	52.3	2376	10	MMSOX1	X94126 M.musculus
27	248.2	49.6	855	4	S69429	S69429 testis-dete
28	237.6	47.5	1824	5	AB011803	AB011803 Gallus ga
29	234.8	47.0	1692	9	AF264713	AF264713 Homo sapi
30	234.8	47.0	2509	9	AX201590	AX201590 Sequence
31	234.8	47.0	2509	9	HSSOX3	X71135 H.sapiens s
c 32	234.8	47.0	148598	9	AL121875	AL121875 Human DNA
33	230.2	46.0	1808	5	GGU12467	U12467 Gallus gall
34	227.6	45.5	664	9	HSA322962	AJ322962 Homo sapi
c 35	225	45.0	8372	6	AX201591	AX201591 Sequence
36	225	45.0	8372	6	AF107044	AF107044 Homo sapi
37	222.2	44.4	1725	10	AY069926	AY069926 Mus muscu
38	221.8	44.4	180538	9	AL137061	AL137061 Human DNA
c 39	220.4	44.1	1786	5	OLA245396	AJ245396 Oryzias 1
40	217.8	43.6	412	6	AX321149	AX321149 Sequence
41	214.8	43.0	1691	9	XLSOX3PRT	Y07542 X.laevis mR
42	211	42.2	2051	9	AB063098	AB063098 Macaca fa
43	209.2	41.8	1161	9	AX001337	AX001337 Sequence
44	208	41.6	957	5	AF193760	AF193760 Gallus ga
45	208	41.6	2315	5	AB026622	AB026622 Gallus ga

ALIGNMENTS

RESULT 1
AX201582

LOCUS   AX201582   1085 bp   DNA   linear   PAT 30-AUG-2001
DEFINITION   Sequence 3 from Patent WO0153349.
ACCESSION   AX201582
VERSION   AX201582.1 GI:15391430
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1085)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
                Chen,Y.T.
  TITLE   Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 3 26-JUL-2001;
                LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
                SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
                INC. (US)
FEATURES   Location/Qualifiers
     source   1..1085
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

```
BASE COUNT      253 a     348 c     348 g     136 t
ORIGIN

Qy      1 CACAGCCCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGGCCCGCAGCAA    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACAGCCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGGCCCGCAGCAA    60

Qy     61 ACTTCGGGGGGCGCGGCGGCCGCAACTCCACCGCGCGGCCGCAACCAGAAAAAC       120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACTTCGGGGGGCGCGGCGGCCGCAACTCCACCGCGCGGCCGCAACACAGAAAAAC      120

Qy    121 AGCCCCGGACACCCGGTCAAGCGGCCCCATGAAATGCCTTCATGGTGTGGCTCCCGGCGCGG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCCCGGACACCCGGTCAAGCGGCCCCATGAAATGCCTTCATGGTGTGGCTCCCGGCGCGG  180

Qy    181 CGCAAGATGGCCCCAGGAGAGAACCCCAAGATGCACAACTCGGAGATCAGAAGCGCCCTGGGC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCAAGATGGCCCCAGGAGAGAACCCCAAGATGCACAACTCGGAGATCAGAAGCGCCCTGGGC  240

Qy    241 GCCGAGTGGAAACTTTTGTCGGAGACGAGAGGCGCGGTTCATCGACGAGGCGTAAGCCG      300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCCGAGTGGAAACTTTTGTCGGAGACGAGAGGCGCGGTTCATCGACGAGGCGTAAGCCG      300

Qy    301 CTGCGAGCGCTGCCACATGAAGGCAGCACCCGGCCCCGGCGGCAACCAGAAAAACC       360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGCGAGCGCTGCCACATGAAGGCAGCACCCGGCCCCGGCGATTATAAAATACCGG      360

Qy    361 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGGTGGGTCGGGGTCGGG      420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGGTGGGTCGGGGTCGGG      420

Qy    421 AATAGCATGGGAGGGGAGGCGGGTGGGGGTCGGGGTCGGGGTCGGGGTCGGGGTCGGG      480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATAGCATGGGCGCGGGGGTCGGGGGTGGGGGGCCTGGGCGTGAACCAGCAGGC       480

Qy    481 ATGGACAGTTACGGCGCACAT     500
          |||||||||||||||||||||
Db    481 ATGGACAGTTACGGCGCACAT    500
```

RESULT 2
LOCUS   HSSOX2G   1085 bp   mRNA   linear   PRI 28-JUL-1995
DEFINITION   H.sapiens sox-2 mRNA (partial).
ACCESSION   Z31560
VERSION   Z31560.1 GI:854181
KEYWORDS   sox-2 gene.
SOURCE   human.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1085)
  AUTHORS   Stevanovic,M., Zuffardi,O., Collignon,J., Lovell-Badge,R. and
               Goodfellow,P.
  TITLE   The cDNA sequence and chromosomal location of the human sox2 gene'
  JOURNAL   Mamm. Genome 5 (10), 640-642 (1994)
  MEDLINE   95152171
REFERENCE   2 (bases 1 to 1085)
  AUTHORS   Stevanovic,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-MAR-1994) Stevanovic M., Institute of Molecular
               Genetics and Genetic Engineering, Vojvode Stepe 283, Belgrade,
               Serbia, Yugoslavia, 11000
FEATURES   Location/Qualifiers
     source   1..1085
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="FBCL1"

BASE COUNT   253 a   348 c   348 g   136 t
ORIGIN

```
Qy      1 CACACGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGGCCCGCAGCAA    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACACGCCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGGCCCGCAGCAA   60

Qy     61 ACTTCGGGGGGCGCGGCGGCCGCAACTCCACCGCGCGGCCGCAACCAGAAAAAC       120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACTTCGGGGGGCGCGGCGGCCGCAACTCCACCGCGCGGCCGCAACACAGAAAAAC      120

Qy    121 AGCCCCGGACACCCGGTCAAGCGGCCCCATGAAATGCCTTCATGGTGTGGCTCCCGGCGCGG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCCCGGACACCCGGTCAAGCGGCCCCATGAAATGCCTTCATGGTGTGGCTCCCGGCGCGG  180

Qy    181 CGCAAGATGGCCCCAGGAGAGAACCCCAAGATGCACAACTCGGAGATCAGAAGCGCCCTGGGC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCAAGATGGCCCCAGGAGAGAACCCCAAGATGCACAACTCGGAGATCAGAAGCGCCCTGGGC  240

Qy    241 GCCGAGTGGAAACTTTTGTCGGAGACGAGAGGCGCGGTTCATCGACGAGGCGTAAGCCG      300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCCGAGTGGAAACTTTTGTCGGAGACGAGAGGCGCGGTTCATCGACGAGGCGTAAGCCG      300

Qy    301 CTGCGAGCGCTGCACATGAAGGAGGATAACTGTAGCACCCGGATTATAAAATACCGGC      360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTGCGAGCGCTGCACATGAAGGAGGATAACTGTAGCACCCGGATTATAAAATACCGGC      360

Qy    361 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGGTGGGTCGGGGTCGGG      420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGGTGGGTCGGGGTCGGG      420

Qy    421 AATAGCATGGGCGCGGGGGTCGGGGGTGGGGGGCCTGGGCGTGAACCAGCAGGC      480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATAGCATGGGCGCGGGGGTCGGGGGTGGGGGGCCTGGGCGTGAACCAGCAGGC      480

Qy    481 ATGGACAGTTACGGCGCACAT     500
          |||||||||||||||||||||
Db    481 ATGGACAGTTACGGCGCACAT    500
```

RESULT 3
LOCUS   HUMHMGBOX   1109 bp   mRNA   linear   PRI 08-MAY-1993
DEFINITION   Homo sapiens (clone 6AR33) HMG box mRNA, 3' end cds.
ACCESSION   L07335
VERSION   L07335.1 GI:184239
KEYWORDS   HMG box.
SOURCE   Homo sapiens retina cDNA to mRNA.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1109)
  AUTHORS   Sadler,L.A., Badzioch,M.D., Wagner,M., Graves,K.A., Swaroop,A.,

FEATURES   Location/Qualifiers
     gene   1..966
                     /gene="sox-2"
     CDS   <1..966
                     /gene="sox-2"
                     /codon_start=1
                     /citation=[1]
                     /product="HSARMYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDR
                     VKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLR
                     ALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQR
                     MDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQ
                     TYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRD
                     MISMYLFPGAEVPEPAAPSRLHMSQHYQSGPVPGTAINGTLPLSHM"
                     /protein_id="CAA84435.1"
                     /db_xref="GI:854182"
                     /db_xref="SWISS-PROT:P48431"
                     /dev_stage="fetus"

```
TITLE       Yang-Feng,T.L., Zheng,K. and Daiger,S.P.
            A novel human retinal cDNA with homology to the SOX gene family
JOURNAL     Unpublished (1992)
FEATURES             Location/Qualifiers
     source          1..1109
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="chromosome 3"
                     /tissue_type="retina"
     CDS             <1..984
                     /note="putative"
                     /codon_start=1
                     /protein_id="AAA35997.1"
                     /db_xref="GI:184240"
                     /translation="RRAGPAHSARMYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQ
                     KNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFID
                     EAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLG
                     AGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMPMHRYDVSALQYN
                     SMTSSQTYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPVVTSSSHSRAPCQ
                     XGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQSGPVPGTRAINGTLPLSHM"
     misc_feature    146..390
                     /standard_name="HMG box"
                     /note="putative"
                     /function="DNA binding"
BASE COUNT      257 a    357 c    357 g    137 t      1 others
ORIGIN

Query Match             100.0%;  Score 500;  DB 9;  Length 1109;
  Best Local Similarity 100.0%;  Pred. No. 1e-74;
  Matches 500;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 CACAGCGCCCGCCCATGTACAACAATGATGGAGACGGAGCTGAAGCCGCCGCCGCAGCCAA 60
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     19 CACAGCGCCCCGCCCATGTACAACAATGATGGAGACGGAGCTGAAGCCGCCGCCGCAGCCA 78

Qy     61 ACTTCGGGGGCGGGGGCGGGGGCGGCAACTCCACCGCGGCGGCGGCCGGGGGC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79 ACTTCGGGGGCGGGGGCGGGGGCGGCAACTCCACCGCGGCGGCGGCCGGGGGC 138

Qy    121 AGCCCGGACCCGGTCAAGCGCCCCATGAATGCCTTCATGGTGTGGTCCCGGGGCCAGCGG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 AGCCCGGACCCGGTCAAGCGCCCCATGAATGCCTTCATGGTGTGGTCCCGGGGCCAGCGG 198

Qy    181 CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC 258

Qy    241 GCCGAGTGGAAACTTTTGTCCGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 GCCGAGTGGAAACTTTTGTCCGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGG 318

Qy    301 CTGCGAGCGCTGCACATGAAGGAGCACCCCGGATTATAAAATACCGGCCCGCCGGGAAACC 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 CTGCGAGCGCTGCACATGAAGGAGCACCCCGGATTATAAAATACCGGCCCGCCGGGAAAACC 378

Qy    361 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGCCTGCTGGCCCCGGGGGGC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGCCTGCTGGCCCCGGGGGGC 438

Qy    421 AATAGCATGGCCAGCGGGGGTCGGGGTGGGGCCGGGCTCGGGGCGGGGGGGAACCAGCGC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 AATAGCATGGCCAGCGGGGGTCGGGGTGGGGCCGGGCTCGGGGCGGGGGGGAACCAGCGC 498

Qy    481 ATGGACAGTTACGCGCACAT 500
          ||||||||||||||||||||
Db    499 ATGGACAGTTACGCGCACAT 518
```

```
RESULT 4
BC013923
LOCUS       BC013923                1181 bp    mRNA    linear   PRI 10-SEP-2001
DEFINITION  Homo sapiens, Similar to SRY-box containing gene 2, clone MGC:2413
            IMAGE:2823424, mRNA, complete cds.
ACCESSION   BC013923
VERSION     BC013923.1  GI:15530269
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1181)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: b Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1..1181
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:2413 IMAGE:2823424"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             85..1038
                     /codon_start=1
                     /product="Similar to SRY-box containing gene 2"
                     /protein_id="AAH13923.1"
                     /db_xref="GI:15530270"
                     /translation="MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKRP
                     MNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHM
                     KEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSY
                     AHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMPMHRYDVSALQYNSMTSSQTYMN
                     GSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPVVTSSSHSRAPCQAGDLRDMISM
                     YLPGAEVPEPAAPSRLHMSQHYQSGPVPGTRAINGTLPLSHM"
BASE COUNT      283 a    379 c    381 g    138 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 9;  Length 1181;
  Best Local Similarity 100.0%;  Pred. No. 1e-74;
  Matches 500;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 CACAGCGCCCGCCCATGTACAACAATGATGGAGACGGAGCTGAAGCCGCCGCCGCAGCCAA 60
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     73 CACAGCGCCCCGCCCATGTACAACAATGATGGAGACGGAGCTGAAGCCGCCGCCGCAGCCA 132

Qy     61 ACTTCGGGGGCGGGGGCGGGGGCGGCAACTCCACCGCGGCGGCGGCCGGGGGC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 ACTTCGGGGGCGGGGGCGGGGGCGGCAACTCCACCGCGGCGGCGGCCGGGGGCCAGCGG 192

Qy    121 AGCCCGGACCCGGTCAAGCGCCCCATGAATGCCTTCATGGTGTGGTCCCGGGGCCAGCGG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 AGCCCGGACCCGGTCAAGCGCCCCATGAATGCCTTCATGGTGTGGTCCCGGGGCCAGCGG 252

Qy    181 CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC 240
```

RESULT 5

LOCUS   HSA327010

LOCUS   HSA327010   832 bp   DNA   linear   PRI 01-OCT-2001
DEFINITION   Homo sapiens genomic sequence surrounding NotI site, clone
               NL1-2D4R.
ACCESSION   AJ327010
VERSION   AJ327010.1 GI:15871428
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 832)
AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
               Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
               Muravenko,O.V., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L.,
               Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
TITLE   Analysis of NotI flanking sequences: a new tool for gene discovery
               and verification of the human genome
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 832)
AUTHORS   Zabarovsky,E.R.
TITLE   Direct Submission
JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
               Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
               Sweden
FEATURES   Location/Qualifiers
     source   1..832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NL1-2D4R"
BASE COUNT   175 a   273 c   269 g   114 t   1 others
ORIGIN

Query Match   99.4%; Score 496.8; DB 9; Length 832;
Best Local Similarity   99.6%;
Matches 498; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 CACAGCGCCCCCATGTACACACATGATGGAGACGTGAAGCCGCCCCCAGCAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 CACACGCGCCCCCATGTACACACATGATGGAGACGTGAAGCCGCCCCCAGCAA 86

Qy   61 ACTTCGGGGGGGGGGGCGGCGGCCAACTCCACCCGCGCCGCCGCGGGAAAAAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87 ACTTCGGGGGGGGGGGCGGCGGCCAACTCCACCCGCGCCGCCGCGGGAAAAAC 146

Qy   121 AGCCCGGACCGCGTCAAGCGCGCCCATGAAATGCCTTCAGTGGTTGGTCCCGCGGGCAGCGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   147 AGCCCGGACCGCGTCAAGCGCGCCATGAAATGCCTTCATGGTGTTGGTCCCGCGCGGGCAGCGG 206

Qy   181 CGCAAGATGGCCCAGGAGAACCCAAGATGCACAACTCGGAGATCACGAGGCCTTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 CGCAAGATGGCCCAGGAGAACCCAAGATGCACAACTCGGAGATCACGAGGCCTTGGGC 312

Qy   241 GCCGAGTGGAAAACTTTTTGTCGGAGACGGAGAAGCCGCCGTTCATCGAGCAGCTAAGCGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 GCCGAGTGGAAAACTTTTTGTCGGAGACGGAGAAGCCGCCGTTCATCGAGCAGCTAAGCGG 372

Qy   301 CTGCGAGCGCGTGCACATGAAGGAGGACGCACCCCGGATTATAAATACCGGCCCCGGGAAACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 CTGCGAGCGCGTGCACATGAAGGAGGACGCACCCCGGATTATAAATACCGGCCCCGGGAAACC 432

Qy   361 AAGACGCTCATGAAGAAGAAGGATAAGTACACGCTGCCCGCGGCGCGGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 AAGACGCTCATGAAGAAGAAGGATAAGTACACGCTGCCCGCGGCGCGGC 492

Qy   421 AATAGCATGGCGAGCGGGGTCGGGGGGGGCGGGGGGGGGGGGGGGGGGGAAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 AATAGCATGGCGAGCGGGGTCGGGGGGGGCGGGGGGGGGGGGGGGGGGGGAAACC 552

Qy   481 ATGGACAGTTACGCGCCACAT 500
              |||||||||||||||||||||
Db   553 ATGGACAGTTACGCGCCACAT 572

RESULT 6

LOCUS   OASOX2GEN

LOCUS   OASOX2GEN   3054 bp   DNA   linear   MAM 21-MAY-1997
DEFINITION   O.aries SOX-2 gene.
ACCESSION   X96997
VERSION   X96997.1 GI:1261960
KEYWORDS   SOX-2; SOX2 protein.
SOURCE   sheep.
ORGANISM   Ovis aries
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
               Bovidae; Caprinae; Ovis.
REFERENCE   1 (bases 1 to 3054)
AUTHORS   Payen,E., Pailhoux,E., Gianquinto,L., Hayes,H., Le Pennec,N.,
               Bezard,J. and Cotinot,C.
TITLE   The ovine SOX2 gene: sequence, chromosomal localization and gonadal
               expression
JOURNAL   Gene 189 (1), 143-147 (1997)
MEDLINE   97305160
REFERENCE   2 (bases 1 to 3054)
AUTHORS   Cotinot,C.Y.
TITLE   Direct Submission
JOURNAL   Submitted (27-MAR-1996) C. Cotinot, Institut National de la
               Recherche Agronomique, Biologie Cellulaire et Moleculaire, Bat.
               Biotechnologies, 78350 Jouy en Josas, FRANCE
FEATURES   Location/Qualifiers
     source   1..3054
                     /organism="ovis aries"
                     /db_xref="taxon:9940"
                     /chromosome="1"
                     /map="q33"
     gene   262..1224
                     /gene="SOX-2"
     CDS   262..1224
                     /gene="SOX-2"
                     /codon_start=1
                     /product="SOX-2 protein"
                     /protein_id="CAA65725.1"
                     /db_xref="GI:1261961"
                     /db_xref="SWISS-PROT:P54231"
                     /translation="MYNMMETELKPPGFQQOTSGGGGGGGGGNSTAAAGGNQKNSPDRV
                     KRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEARLRA
                     LHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRM
                     DSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQT
                     YMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPVVTSSSHSRAPCQAGDLRDM
                     ISMYLPGAEVPEPAAPSRLHMSQHYQSGPVPGTRAINGTLPLSHM"
BASE COUNT   830 a   741 c   744 g   739 t

Query Match   99.0%; Score 496.8; DB 9; Length 3054;
Best Local Similarity   99.6%;
Matches 498; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   181 CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCACGAGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   207 CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCACGAGGCC 266

Qy   241 GCCGAGTGGAAAACTTTTTGTCGGAGACGGAGAAGCCGCCGTTCATCGAGCAGCTAAGCGG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   267 GCCGAGTGGAAAACTTTTTGTCGGAGACGGAGAAGCCGCCGTTCATCGAGCAGCTAAGCGG 326

Qy   301 CTGCGAGCGCGTGCACATGAAGGAGGACGCACCCCGGATTATAAATACCGGCCCCGGGAAACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   327 CTGCGAGCGCGTGCACATGAAGGAGGACGCACCCCGGATTATAAATACCGGCCCCGGGAAACC 386

Qy   361 AAGACGCTCATGAAGAAGAAGGATAAGTACACGCTGCCCGCGGCGCTGCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   387 AAGACGCTCATGAAGAAGAAGGATAAGTACACGCTGCCCGCGGCGCTGCTG 446

Qy   421 AATAGCATGGCGAGCGGGGTCGGGGGGGGCGGGGGGGGGGGGGGGGGGGAAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   447 AATAGCATGGCGAGCGGGGTCGGGGGGGGCGGGGGGGGGGGGGGGGGGGGAAACC 506

Qy   481 ATGGACAGTTACGCGCCACAT 500
              |||||||||||||||||||||
Db   507 ATGGACAGTTACGCGCCACAT 526

```
ORIGIN
Best Local Similarity  94.3%;  Score 449;  DB 4;  Length 3054;
Query Match  89.8%;  Pred. No. 3.1e-66;
Matches 480;  Conservative  0;  Mismatches 20;  Indels 9;  Gaps 1;

Qy     1  CACAGCGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCGCCCGGGCCCGCAGCAA   60
          |||||  ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   250  CACAGCACCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCGCCCGGGCCCGCAGCAA  309

Qy    61  ACTTCGGGGGGCGGCGGG------------CGGCAACTCCACCCGGGGGGGGGGGGGCAAC  120
          |||||||||||||||||||            |||||||| ||||||||||||||||||||||
Db   310  ACTTCGGGGGGCGGCGGGGGGGGGGGGGGCGGCGGCAACTCCACCCGGGGGGGGGGGGCAAC  369

Qy   112  CAGAAAAACAGCCCGGAACCGGGCGTCAAGCGCCCATGAATGCCTTCATCATGGTGTCCGC  171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   370  CAGAAGAACAGCCCGGAACCGGGCGTCAAGCGCCCATGAATGCCTTCATCATGGTGTCCGC  429

Qy   172  GGGCAGCGGCGGGCCAAGATGGGCCCAGGAGAACCCCAGAAGCACAACTCGGAGATCACGA  231
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   430  GGGCAGCGGCGGGCCAAGATGGGCCCAGGAGAACCCCAGAAGCACAACTCGGAGATCACGA  489

Qy   232  CGCCTGGGGCCCCGAGTGGGAAAACTTTTTGTCCGGGAGACGGGAGAAGCGGCCGTTCATC  291
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   490  CGCTTGGGGCCCCGAGTGGGAAAACTTTTTGTCGGGAGACGGGAGAAGCGGCCGTTCATC  549

Qy   292  GCTAAGCGCGCTGCGAGCGCTGCACATGAAGGAGCACACCCGGATTATAAATACCGGCCCG  351
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   550  GCCAAGCGCGCTGCGAGCGCTGCACATGAAGGAGCACACCCGGATTATAAATACCGGCCCG  609

Qy   352  CGGAAAACCAAGAGACGCTCATGAAGAAGGAATAAGTACACACGCTGCCTGCCTGGCC  411
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   610  CGGAAAACCAAGAGACGCTCATGAAGAAGGAATAAGTACACACGCTGCCTGCCTGGCC  669

Qy   412  CCCGGGGCGGCAAATAGCATGGGCGGGGTCGGGGGTGGGGGGCGCCTGGGGCCGGGGCTG  471
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   670  CCGGGGGCGGCAAACAGCCATGGGCGGGGTCGGGGGTGGGGGGCGCCTGGGGCCGGGGCTG  729

Qy   472  AACCAGCGCCATGGACAGTTACGCGCACAT  500
          |||||||||||||||||||||||||||
Db   730  AACCAGCGCCATGGACAGCTACGCGCACAT  758
```

```
RESULT 7
LOCUS       HSA324649        749 bp    DNA     linear   PRI 01-OCT-2001
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NR5-FN13R.
ACCESSION   AJ324649
VERSION     AJ324649.1  GI:15869028
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
            Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
            Muravenko,O.V., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L.,
            Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     Analysis of NotI flanking sequences: a new tool for gene discovery
            and verification of the human genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 749)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
BASE COUNT      161 a     231 c     245 g     110 t     2 others
ORIGIN                                                    /clone="NR5-FN13R"
Best Local Similarity  89.6%;  Score 447.8;  DB 9;  Length 749;
Query Match  95.0%;  Pred. No. 6.2e-66;
Matches 472;  Conservative  0;  Mismatches 24;  Indels 1;  Gaps 1;

Qy     1  CACAGCGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCGCCCGGGCCCGCAGCAA   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    27  CACAGCGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCGCCCGGGCCCGCAGCAA   86

Qy    61  ACTTCGGGGGGCGGCGGGGGGGGGGGCGGCAACTCCACCCGGGGGGGGGGGGGCAACAGAAAAAC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    87  ACTTCGGGGGGCGGCGGGGGGGGGGGCGGCAACTCCACCCGGGGGGGGGGGGGCAACAGAAAAAC  146

Qy   121  AGCCCGGAACCGGGCGTCAAGCGCCCATGAATGCCTTCATCATGGTGTCCGCGGGCAGCGG  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   147  AGCCCGGAACCGGGCGTCAAGCGCCCATGAATGCCTTCATCATGGTGTCCGCGGGCAGCGG  206

Qy   181  CGCAAGGATGGGCCCAGGAGAACCCCAGAAGCACAACTCGGAGATCACGACGCCTGGGGC  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   207  CGCAAGGATGGGCCCAGGAGAACCCCAGAAGCACAACTCGGAGATCACGACGCCTGGGGC  266

Qy   241  GCCGAGTGGGAAAACTTTTTGTCCGGAGACGGGAGAAGCGGCCGTTCATCGACGAGGGTTAAGCGG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   267  GCCGAGTGGGAAAACTTTTTGTCGGGAGACGGGAGAAGCGGCCGTTCATCGACGAGGGTTAAGCGG  326

Qy   301  CTGCGAGCGCTGCACATGAAGGAGCACACCCGGATTATAAATACCGGCCCGCGGAAAACC  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   327  CTGCGAGCGCTGCACATGAAGGAGCACACCCGGATTATAAATACCGGCCCGCGGAAAACC  386

Qy   361  AAGACGCTCATGAAGAAGGAATAAGTACACACGCTGCCTGCCTGGCCCCCGGGGCGGC  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   387  AAGACGCTCATGAAGAAGGAATAAGTACACACGCTGCCTGCCTGGCCCCCGGGGCGGC  446

Qy   421  AATAGCATGGGCGGGGTCGGGGGTGGGGGGCGCCTGGGGCCGGGGCTGAACCAGCGCC  480
          |||  ||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   447  AATATCATGGCAAGCGGGNGGTCGGGTGGGGGGCCCCG-CTTGGGCCCGGGGAACANCGCA  505

Qy   481  ATGGACAGTTACGCGCCA  497
          |||  |||||||||||
Db   506  CTGGCAAGTCCCCCCCA  522
```

```
RESULT 8
LOCUS       AL606746       203344 bp    DNA     linear   HTG 30-NOV-2001
DEFINITION  Mus musculus chromosome 3 clone RP23-423J10, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL606746
VERSION     AL606746.16  GI:17381397
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 5, 2001 this sequence version replaced gi:17128816.
            -------------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------------- Project Information
            Center project name: bM423J10
            --------------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
```

```
Sequencing vector: plasmid; L08752; 100% of reads
    Chemistry: Dye-terminator Big Dye; 0% of reads
    Chemistry: Dye-primer Big Dye; 99% of reads
    Consensus quality: 203311 bases at least Q40
    Consensus quality: 203311 bases at least Q30
    Consensus quality: 203311 bases at least Q20
    Consensus quality: 203343 bases at least Q20
    Consensus quality: 203344 bases at least Q20
    Consensus quality: 203344 sum-of-contigs
    Insert size: 203344 sum-of-contigs
    Insert size: 200969; 7.9% error; agarose-fp
    Quality coverage: 11.99x in Q20 bases; sum-of-contigs Quality
    coverage: 12.13x in Q20 bases; agarose-fp
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.

FEATURES             Location/Qualifiers
     source          1..203344
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="3"
                     /clone="RP23-423J10"
                     /clone_lib="RPCI-23"
     misc_feature    1..203344
                     /note="assembly_fragment:01841"

BASE COUNT    55075 a   45698 c   45713 g   56858 t
ORIGIN

Query Match          87.8%;  Score 439.2;  DB 2;  Length 203344;
Best Local Similarity  93.3%;  Pred. No. 7.2e-65;
Matches  472;  Conservative  0;  Mismatches  28;  Indels  6;  Gaps  1;

Qy      1 CACAGCGGCCCGCCGCCATGTACACACATGATGGAGACGGAGCTGAAGCCGCCCCGGCCCAGCAA   60
          |||||| |||||||||| || |||||||||||||||||||||||||||||||| ||||||||||
Db  72956 CCCAGCGCCCCGCCGCCCATGTATAACACATGATTGGAGACGGAGCTGAAGCCGCCCCCAGCAA   73015

Qy     61 ACTTCGGGGGGCGGCGC------GGCAACTCCACCGCCGCCGGCGGCGCGGCGCGCAACCAG   114
          ||||||||||||||||||       ||||||||||||||||||||||||||||||||||||
Db  73016 GCTTCGGGGGGCGGGGCGGGGGGGGCAACTCCACCGCCGCCGGCGGCGCGGCGCGCAACCAG   73075

Qy    115 AAAAACAGCCCGGACCGGCGCGTCAAGCGCGCCCATGAATGCCTTCATGGGTGGTCCCGCGGG   174
          ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db  73076 AAGAACAGCCCGGACCGGCGTCAAGAGCGCCCATGAAGCCCTTCATGGTGTATGGTCCCGGGGG   73135

Qy    175 CAGCGGCGCGAAGAGATGGCCCCAGGAGAACCCCAAGAATGCACAACTCGGAGATCAGCAAGCGC   234
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  73136 CAGCGGCGTAAGGATGGCCCAGGAGAACCCCAAGAATGCACAACTCGGAGATCAGCAAGCGC   73195

Qy    235 CTGGGCGCGCCGAGTGGAAAACTTTTGTCGGAGACAGCGAGAAGCCGCGTTCATCGACGAGGCT   294
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  73196 CTGGGCGCGCCGAGTGGAAACTTTTGTCCGAGACGAGAAGCCGCGTTCATCGACGAGGCC   73255

Qy    295 AAGCGGCGTCGCAGCGCTGCACATGAAGGACAAGTATACACTGCCGGGTCTGCTGGCCCCGCCG   354
          |||||||||||| || |||||||||||||||||||||||||||||||||||||||||||||||
Db  73256 AAGCGGCTCGCGCGCTCTGCACATGAAGGACAAGTATACACCGCCGGGTCTGCTGGCCCCGGCCG   73315

Qy    355 AAAAACCAGAGACGCTCATGAAGAAGGATAAGTACACGCCGCCCGCCGCCGCCCCGCCAACCAG   414
          |||||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||||
Db  73316 AAAAACCAAGAGACGCTCATGAAGAAGGATAAGTACACGCCGCCCGGCGGCGCGGCGCGCAACCAG   73375

Qy    415 GGGCGGGCAATAGCATGGCGGACGGGGGGTCGGGGGTGGGCGCCGGCGTGAAC   474
          |||||||||||||||||||||||||||| ||||||||||||||| |||||||||
Db  73376 GGGCGGGAACAGCATGGCGGACGGGGGTTGGGGGTGGGCGCCGGCGTGAAC   73435

Qy    475 CAGGGCCATGGACACTTACCGCGCACAT   500
          |||||||||||| |||||||||||||||
Db  73436 CACGGCCATGGACAGCTACCGCGCACAT   73461
```

```
RESULT 9
MMSOX2
LOCUS       MMSOX2                  2418 bp    DNA       linear   ROD 13-MAR-1996
DEFINITION  M.musculus SOX2 gene.
ACCESSION   X94127
VERSION     X94127.1  GI:1209429
KEYWORDS    sox-2 gene; SOX2 protein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2418)
  AUTHORS   Collignon,J., Sockananthan,S., Hacker,A., Cohen-Tannoudji,M.,
            Norris,D., Rastan,S., Stevanovic,M., Goodfellow,P.N. and
            Lovell-Badge,R.
  TITLE     A comparison of the properties of Sox-3 with Sry and two related
            genes, Sox-1 and Sox-2
  JOURNAL   Development 122 (2), 509-520 (1996)
  MEDLINE   96189340
REFERENCE   2  (bases 1 to 2418)
  AUTHORS   Sockananthan,T.E.L.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1995) TEL.S. Sockananthan, National Institute For
            Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
COMMENT     Overlaps with X55491, D50603 and U31967.
FEATURES             Location/Qualifiers
     source          1..2418
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /dev_stage="8.5 dpc"
     gene            407..1366
                     /gene="sox-2"
     CDS             407..1366
                     /gene="sox-2"
                     /function="transcription factor"
                     /codon_start=1
                     /product="SOX2 protein"
                     /protein_id="CAA63847.1"
                     /db_xref="GI:1209430"
                     /db_xref="MGD:MGI:98364"
                     /db_xref="SWISS-PROT:P48432"
                     /translation="MYNMMETELKPGPQQASGGGGGGGNATAAATGNQKNSPDRVK
                     RPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRAL
                     HMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMD
                     SYAHMNGWSNGSYSMMQEOLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTY
                     MNGSPTYSMSYSQOGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMI
                     SMYLPGAEVPEPAAPSRLHMAQHYQSGPVPGTAINGTLPLSHM"
     misc_feature    527..763
                     /gene="sox-2"
                     /note="HMG box"

BASE COUNT    645 a    602 c    653 g    518 t
ORIGIN

Query Match          87.2%;  Score 436;  DB 10;  Length 2418;
Best Local Similarity  92.9%;  Pred. No. 4.9e-64;
Matches  470;  Conservative  0;  Mismatches  30;  Indels  6;  Gaps  1;

Qy      1 CACAGCGGCCCGCCGCCATGTACACACATGATGGAGACGGAGCTGAAGCCGCCCCGGCCCAGCAA   60
          |||||| |||||||||| || |||||||||||||||||||||||||||||||| ||||||||||
Db    395 CCCAGCGCGCCGCCGCCCATGTATAACACATGATTGGAGACGGAGCTGAAGCCGCCGCCCCGGCAA   454

Qy     61 ACTTCGGGGGGCGGCGC------GGCAACTCCACCGCCGCCGGCGGCGCGGCGCGCAACCAG   114
          ||||||||||||||||||       ||||||||||||||||||||||||||||||||||||
Db    455 GCTTCGGGGGGCGGGGCGGGGGGGGCAACTCCACCGCCGCCGGCGGCGCGGCGCGCAACCAG   514

Qy    115 AAAAACAGCCCGGACCGGCGCGTCAAGCGCGCCCATGAATGCCTTCATGGGTGGTCCCGCGGG   174
          ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db    515 AAGAACAGCCCGGACCGGCGTCAAGAGCGCCCATGAAGCCCTTCATGGTGTATGGTCCCGGGGG   574

Qy    175 CAGCGGCGCGAAGAGATGGCCCCAGGAGAACCCCAAGAATGCACAACTCGGAGATCAGCAAGCGC   234
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    575 CAGCGGCGTAAGGATGGCCCAGGAGAACCCCAAGAATGCACAACTCGGAGATCAGCAAGCGC   634

Qy    235 CTGGGCGCGCCGAGTGGAAAACTTTTGTCGGAGACAGCGAGAAGCCGCGTTCATCGACGAGGCT   294
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    635 CTGGGCGCGCCGAGTGGAAACTTTTGTCCGAGACGAGAAGCCGCGTTCATCGACGAGGCC   694
```

RESULT 10

LOCUS   AX305428   2283 bp   DNA   linear   PAT 11-DEC-2001
DEFINITION   Sequence 179 from Patent WO0188188.
ACCESSION   AX305428
VERSION   AX305428.1   GI:17644960
KEYWORDS
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
  AUTHORS   Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
  TITLE   Method for examining ischemic conditions
  JOURNAL   Patent: WO 0188188-A 179 22-NOV-2001;
             School Juridical Person Nihon University (JP)
FEATURES   Location/Qualifiers
     source   1..2283
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"

BASE COUNT   597 a   583 c   614 g   488 t   1 others
ORIGIN

Query Match   86.6%;   Score 432.8;   DB 6;   Length 2283;
Best Local Similarity 92.5%;   Pred. No. 1.7e-63;
Matches 468; Conservative 0; Mismatches 32; Indels 6; Gaps 1;

```
Qy    295 AAGCGGCTGCGACGGCGCCGCCCCCGGG ATTATAAAATACCGCCGCGCGG 354
          |||||||||||||||||||||||||||| ||||||||||||||||||| |||
Db    695 AAGCGGCTGCGCGCTCTGCACATGAAAGA ACCACCCGGATTATAAAATACCGCCGGC 754

Qy    355 AAAACCAAGACGCTCATGAAGAGAAGGATAAGTACACGCCCGGATTATATAAATACCGCCCC 414
          ||||||||||||||||||||| ||||||||||||||| |||||||  |||||||||||||||
Db    755 AAAACCAAGACGCTCATGAAGAAGGATAAGTACACGCCGCTTCCCGGAGGCTTGCTGGCCCC 814

Qy    415 GGCGGGCAATAGCATGGCGAGCGGGGTCGGGGGTGTGGGGGCGCCCGGGCGCGGGCGTGAAC 474
          |||| |||||||||||||||||||||||||||||||||||| |||||||  ||||||| |||
Db    815 GGCGGGGAACAGCATGGCCGAGCGGGGTTGGGGGTGCGGGGCGCCCGTGGGTGCGGGGGTGAAC 874

Qy    475 CAGCGCCATGGACAGTTACGCGCCACAT 500
          ||||||||||||||||||||||||||||
Db    875 CAGCGGCATGGACACCTACCGCGCACAT 900
```

RESULT 11

LOCUS   MMU31967   2283 bp   mRNA   linear   ROD 14-AUG-1998
DEFINITION   Mus musculus high mobility group box protein (sox2) mRNA, complete
             cds.
ACCESSION   U31967
VERSION   U31967.1   GI:3419872
KEYWORDS
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2283)
  AUTHORS   Yuan,H., Corbi,N., Basilico,C. and Dailey,L.
  TITLE   Developmental-specific activity of the FGF-4 enhancer requires the
             synergistic action of Sox2 and Oct-3
  JOURNAL   Genes Dev. 9 (21), 2635-2645 (1995)
  MEDLINE   96067156
REFERENCE   2 (bases 1 to 2283)
  AUTHORS   Basilico,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-JUL-1995) Claudio Basilico, Microbiology, NYU Medical
             Center, 550 First Avenue, New York, NY 10016, USA
COMMENT   On Aug 14, 1998 this sequence version replaced gi:1101763.
FEATURES   Location/Qualifiers
     source   1..2283
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene   1..2283
                     /gene="Sox2"
     CDS   362..1321
                     /gene="Sox2"
                     /codon_start=1
                     /product="high mobility group box protein SOX2"
                     /protein_id="AAC31791.1"
                     /db_xref="GI:3419873"
                     /translation="MYNMMETELKPPGPQQASGGGGGGGGGNATAAATGGNQKNSPDRVK
                     RPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEARLRAL
                     HMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGGGLNQRMD
                     SYAHMNGWSNGSYSMMQEQLGYPQHPGLNAHGAAQMQPMHRYVVSALQYNSMTSSQTY
                     MNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMI
                     SMYLPGAEVPEPAAPSRLHMAQHYQSGPVPGTAKYGTLPLSHM"
     polyA_site   2283
                     /gene="Sox2"
                     /note="24 A nucleotides"

BASE COUNT   597 a   583 c   614 g   488 t   1 others
ORIGIN

Query Match   86.6%;   Score 432.8;   DB 10;   Length 2283;
Best Local Similarity 92.5%;   Pred. No. 1.7e-63;
Matches 468; Conservative 0; Mismatches 32; Indels 6; Gaps 1;

```
Qy      1 CACACGCGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGGCCCGCAGCAA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 CCCAGCGCCCCGCCATGTATAACATGATGGAAGACGGAGCTGAAGCCGCCGGGGCCCCAACAG 409

Qy     61 ACTTCGGGGGGCGGGGGC------GGCAACTCCACCGCGCGGGGGGGGGGCCGGCCGCAACGC 114
          ||||||||||||||||||      |||||||||||||||||| ||||||||||||||
Db    410 GCTTCGGGGGCGCGCTGCGCGGGGAGGAGGCAACGCCACGGCGCGGGGGGCCGGCCGCAACCAG 469

Qy    115 AAAAACAACGCCCGACCCGCGTCAAGCGGCCCATGAATCCTTCATGGTGTGGTCCCGGCGGG 174
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    470 AAGAACAGCCCGACCCGCGTCAAGAGAGCCCATGAATGCCTTCATGGTGTGGTCCCGGGGG 529

Qy    175 CAGCGCGCGCAAGATGGCCCCAGCGGGGAAGACCCCAAGATGCACAACTCGGAGATCAGCAAGCGC 234
          |||||||||||||||||||||| ||||||||||| |||||||||||||||||||||||| |||
Db    530 CAGCGGCGTAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGC 589

Qy    235 CTGGGCGGCCGAGTGGAAAACTTTTTGTCGGAGACGGAGACGAAGCCGCCGTTCATCGACGAGGCT 294
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 CTGGGCGGCCGAGTGGAAAACTTTTTGTCCGAGACGGAGACGAAGCCGCCGTTCATCGACGAGGCC 649

Qy    295 AAGCGGCTGGCGAGCGCTGCACATGAAGGAGCACCCGGATTATAAATACCGGCCCGGCCCCGGCGG 354
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    650 AAGCGGCTGGCGCTGCACATGAAGGAGCACCCGGATTATAAATACCGGCCCGGCGG 709

Qy    355 AAAAACCAAGACGCTCATGAAGAAGGATAAGTACACGCTTCCCGGAGGCTTGCTGGCCCC 414
          ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    710 AAAAACCAAGACGCTCATGAAGAAGGATAAGTACACGCTTCCCGGAGGCTTGCTGGCCCC 769

Qy    415 GGGCGGCAATAGCATGGCGAGCGGGGTCGGGGGTGGGGGCGCGGGCCTGGGCGGGGGTGAAC 474
          ||||||||||||||||||||||||||||||||||||||| ||||||||||| ||||||||||
Db    770 GGGCGGGGAACAGCAGCATGGCGAGCGGGGGGTTGGGGGTGGGGGGCGGGGGCCTGGGTGGGGGTGAAC 829

Qy    475 CAGCGCCATGGACAGTTACGCGCCACAT 500
          ||||||||||||||||||||||||||||
Db    830 CAGCGGCATGGACACCTACCGCGCACAT 855
```

RESULT 12

LOCUS   CHKSOX2   1186 bp   mRNA   linear   VRT 10-FEB-1999
DEFINITION Chicken mRNA for SOX-2, complete cds.
ACCESSION   D50603
VERSION   D50603.1 GI:849043
KEYWORDS   SOX-2.
SOURCE   Gallus gallus (strain White leghorn) embryo brain (library: lambda
            gt10) cDNA to mRNA, clone 172-B1.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 1186)
  AUTHORS   Kamachi,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-MAY-1995) Yusuke Kamachi, Osaka University, Institute
            for Molecular Cellular Biology; 1-3 Yamadaoka, Suita, Osaka 565,
            Japan (E-mail:i6459ia@center.osaka-u.ac.jp, Tel:06-879-7964,
            Fax:06-877-1738)
REFERENCE   2 (bases 1 to 1186)
  AUTHORS   Kamachi,Y., Sockanathan,S., Liu,Q., Breitman,M., Lovell-Badge,R.
            and Kondoh,H.
  TITLE   Involvement of Sox proteins in lens-specific activation of
            crystallin genes
  JOURNAL   EMBO J. 14 (14), 3510-3519 (1995)
  MEDLINE   95354669
FEATURES   Location/Qualifiers
     source   1..1186
                     /organism="Gallus gallus"
                     /strain="White leghorn"
                     /db_xref="taxon:9031"
                     /tissue_type="brain"
                     /clone_lib="lambda gt10"
                     /dev_stage="embryo"
                     /clone="172-B1"
     CDS   74..1012
                     /function="SRY-type HMG-domain containing transcription
                     factor"
                     /codon_start=1
                     /product="SOX-2"
                     /protein_id="BAA09168.1"
                     /db_xref="GI:849044"
                     /translation="MMETELKPPAPQQTSGGTGNSNAANNQKNSPDRVKRPMNAFM
                     VWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSEAEKRPFIDEAKRLRALHMKEHPD
                     YKYRPRRKTKTLMKKDKYTLPGGLLAPGTNTMTTGVGVGATLGAGVNQRMDSYAHMNG
                     WTNGYGYMMQEEQLGYPQHPGLNAHNNAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTY
                     SMSYSQQGTPGMALGSMGSVVKTESSSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGA"

BASE COUNT   300 a   336 c   343 g   207 t
ORIGIN

                     EVPEPAAPSRLIHMSQHYQSAPVPGTAINGTLPLSHM"

Query Match   71.8%;   Score 359.2;   DB 5;   Length 1186;
Best Local Similarity   84.9%;   Pred. No. 4.2e-51;
Matches   416;   Conservative   0;   Mismatches   68;   Indels   6;   Gaps   1;

Qy   175 CACGCGGCCAAGATGGCCCAGGAGAGACCCCAAGATCAGATGCACAACTCGGAGATCAGCAAGCGC 234
          |||||||| |||||||||||||| |||||||||| ||||||||||||||||||||||||
Db   530 CACGCGGCCTAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGC 589

Qy   235 CTGGGCGCCGAGTGGAAACTTTTGTCGGAGACGAGGAGAAGCGGCCGTTCATCGACGAGGCT 294
          |||||||||||||||||||||| ||||||||||||||||||||||||||||| |||
Db   590 CTGGGCGCCGAGTGGAAACTTTTGTCCGAGACCGAGAAGCGGCCGTTCATCGACGAGGCC 649

Qy   295 AAGCGGCTGCGAGCGCTGCACATGAAAGAGCACCCGGATTATAAATACCGGCCGCGGCGG 354
          |||||||||||| | |||||||||||||| |||||| || |||||||||||||||||
Db   650 AAGCGGCTGCGCGCTCTGCACATGAAGGAGCACCCCGATTATAAATATCGGCCGCGGCGG 709

Qy   355 AAAACCAAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGCCTGCTGGCCCCC 414
          |||||||||| | |||||| ||||||||||||||||| |||||||||| |||||||||||
Db   710 AAAACCAAGACGCTCATGAAGAAGGATAAGTACACGCTTCCCGGCGGCCTGCTGGCCCCC 769

Qy   415 GGCGGCAATAGCATGGCGAGCGGGGTCGGGGTGGGCGCCGGCCTGGGTGCTGGCGTGAAC 474
          ||||||||| |||||||||||||||| ||||||||| || |||||| ||| |||||||||
Db   770 GGCGGGAACAGCATGGCGAGCGGGGTTGGGGTGGGCGCCGGCCTTGCTGGCGTGAAC 829

Qy   475 CAGCGCATGGACAGTTACGCGCACAT 500
          ||||| |||||||| ||||||||||||
Db   830 CAGCGCATGGACAGCTACGCGCACAT 855

RESULT 13

LOCUS   GGU12532   1355 bp   mRNA   linear   VRT 03-OCT-1996
DEFINITION Gallus gallus Sox2 transcription factor (Sox2) mRNA, complete cds.
ACCESSION   U12532
VERSION   U12532.1 GI:595494
KEYWORDS   .
SOURCE   chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 1355)
  AUTHORS   Uwanogho,D., Rex,M., Cartwright,E.J., Pearl,G., Healy,C.,
            Scotting,P.J. and Sharpe,P.T.
  TITLE   Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes
            suggests an interactive role in neuronal development
  JOURNAL   Mech. Dev. 49 (1-2), 23-36 (1995)
  MEDLINE   95267693
REFERENCE   2 (bases 1 to 1355)
  AUTHORS   Healy,C.P.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-JUL-1994) Christopher P. Healy, Craniofacial
            Development, Dental School, UMDS, Guy's Hospital, London, SE1 9RT,
            UK
FEATURES   Location/Qualifiers
     source   1..1355
                     /organism="Gallus gallus"

Qy   237 CCCAGGAGAACCCGAAGATGCACAACTCGGAGATCAGCAAGCGCTCGGGCGCCGAGTGGA 296
          |||||||||||| |||||||||||||||||||||| |||||| |||||| |||||||
Db   297 AACTTTTTATCCGAGGACGGAGAAGAGACCCTTCATTGACGACGAGGCTCCCGAGCTC 356

Qy   251 AACTTTTTGTCGGAGACGAGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGGCTGCGAGCGC 310
          |||||||||||||||||||||||||||||||||||||||||||||||||||| | ||||
Db   297 AACTTTTTATCCGAGGACGGAGAAGAGACCCTTCATTGACGACGAGGCTCCCGAGCTC 356

Qy   311 TGCACATGAAAGGAGCACCCGGATTATAAATACCGGCCGCGGCGGAAAACCAAGACGCTCA 370
          ||||||||| | |||||||||||||||||||||||||||||| ||| ||||||||
Db   357 TGCACATGAAAGGAGCACCCGGATTATAAATACCGGCCGCGGCGGAAAACCAAGACGCTGA 416

Qy   371 TGAAGAAGGATAAGTACACGCTGCCCGGCGGCCTGCTGGCCCCCGGCGGCAATAGCATGG 430
          ||||||||||||| |||||||| |||||||||| ||| |||||||||||| |||||||||
Db   417 TGAAGAAGGATAAGTAACGACTTGCCAGGCGGCCTGCTGGCCCCGCCACCAATACCATGA 476

Qy   431 CGAGCGGGGTCGGGGTGGGCGCCGGCCTGGGTGCTGGCGTGAACCAGCGCATGGACAGTT 490
          |||||||||||||| || ||||| || ||| ||||||| |||||||||| |||||||
Db   477 CGACCGGGGTAGGGGTTGGGGCGCCACCTTGGGTGAACCAGAGGACAGTT 536

Qy   491 ACGCGCACAT 500
          ||||||||||
Db   537 ACGCGCACAT 546

```
5'UTR           1..258
                /db_xref="taxon:9031"
                /gene="Sox2"
gene            1..1355
                /gene="Sox2"
CDS             259..1206
                /gene="Sox2"
                /codon_start=1
                /product="Sox2 transcription factor"
                /protein_id="AAB09662.1"
                /db_xref="GI:595495"
                /translation="MYNMMETELKPPAPQQTSGGTGTGNSNSNAANNQKNSPDRVKRPMN
                AFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSEAEKRPFIDEAKRLRALHMKE
                HPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGTNTMTTGVGVGATLGAGVNQRMDSYAH
                MNGWTNGGYGMMQEQLGYPQHPGLNAHNNAAQMQPMHRYDVSALQYNSMTSSQTYMNGS
                PTYSMSYSQQTPGMALGSMGSVVKTSSSSPPVVTSSSHSRAPCQAGDLRDMISMYL
                PGAEVPEPAAPSRLHMSQHYQSAPVPGTAINGTLPLSHM"
3'UTR           1207..1355
                /gene="Sox2"
misc_feature    367..603
                /gene="Sox2"
                /note="encodes sry-like HMG-box"
BASE COUNT      362 a      373 c      370 g      250 t
ORIGIN

Query Match     71.8%;  Score 359.2;  DB 5;  Length 1355;
Best Local Similarity   84.9%;  Pred. No. 4.1e-51;
Matches 416;  Conservative   0;  Mismatches  68;  Indels   6;  Gaps  1;

Qy     11 GCAATGTACAACAATGATGGAGACGGAGCTGAAGCCGCCAGCCCCACAGCAAACTTCGGGG 70
          ||||||| ||| |||||||||||||| || || | |||||||||||| | || ||||||||
Db    257 GAATGTACAACAATGATGGAGAAACCGAGCTGAAAACCTCACGCCCCGGGGG 316

Qy     71 GCGGGCGGGCGGAACTCACCGGCGGCGGCGGCGGCGGCGGCAGAAAACAGCCCGGACC 130
          || ||||| || |||||| ||| || |||||||||||||||||  || ||| |||||||
Db    317 GAGGCACAGCAAGCTCCAACTCCGCCGGC----- AACAACCAGAGACAGCCCGGACC 370

Qy    131 GCGTCAAGCGGCCCATGAATGCCTTCATGGTGTGGTCCCGCGGCCGCCAAGATGG 190
          |||||||||| |||| ||||| |||| |  || |||||||||||| || | |||||
Db    371 GGGTCAAGCGGCCCATGAACGCCTTCATGGTGTGGTCCCGCGGCCGCGAAAATGG 430

Qy    191 CCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGCGCCGAGTGGA 250
          ||||| || |||| ||||||||||| ||||||||||| |||||||| |||||||||||||
Db    431 CCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAACGCCTCGGCGGCGAGTGGA 490

Qy    251 AACTTTTGTCGAGAGACGGAAGCGGCCGTCCGTGGCGGCCGTGCCGGAGCGC 310
          |||||||| || || | || |||||||| ||| || ||| |||||||||||
Db    491 AACTTTTATCCGAGGCGGAAGAGACCCTTCATTGACGAAGCCAAACGGCTCCGAGCTC 550

Qy    311 TGCACATGAAGGAGCACCCGGATTATAAATACCGGCCGCGGAAAACCAAGAGCGCTCA 370
          ||||||||||||||||||||||| || |||||||| | ||| |||| |||| || || 
Db    551 TGCACATGAAGGAGCACCCGGATTATAAATACCGACCCCGGAGAAAACCAAGACCCTGA 610

Qy    371 TGAAGAAGGATAAGTACACGCTGCCCGGCCGCTGCTGGGCCCGCCCCCGCCGGAATAGCGG 430
          || ||||| ||||| || ||| | ||||| ||| |||||| |||| ||||| || ||||||
Db    611 TGAAGAAGGATAAGTACACGTTCCCAGGGCGCTTACTGGCCCCGGGCCACCAATACCATGA 670

Qy    431 CGAGCGCGGGGTCGGGGTCGGGGTCGGGGTCTGAACCAGCCGCCATGGACAGTT 490
          ||||||||| || ||||| ||||| |||||| | ||||||||||| ||||||||
Db    671 CGACCGGCGGGGTAGGGGTTGGGGTCCACCTTTGGGGAGCTGGGGTGAACCAGGATGGACAGTT 730

Qy    491 ACGCGCACAT 500
          ||||||||||
Db    731 ACGCGCACAT 740
```

```
RESULT 14
LOCUS       AX321190        395 bp    DNA        linear    PAT 15-DEC-2001
DEFINITION  Sequence 207 from Patent WO0177168.
ACCESSION   AX321190
VERSION     AX321190.1 GI:17904744
KEYWORDS    .
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0177168-A 207 18-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..395
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      83 a      128 c      144 g      40 t
ORIGIN

Query Match     70.0%;  Score 350;  DB 6;  Length 395;
Best Local Similarity  100.0%;  Pred. No. 1.7e-49;
Matches 350;  Conservative   0;  Mismatches  0;  Indels   0;  Gaps  0;

Qy      1 CACAGCGCGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGCCGAGCAA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db     46 CACAGCGCGCCCGCCATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGCCGAGCAA 105

Qy     61 ACTTCGGGGGGGGGGGGCGGGCGGCGGCCAACTCCACCGGCGGCGGCGGCCAACCAGAAAAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    106 ACTTCGGGGGGGGGGGGCGGGCGGCGGCCAACTCCACCGGCGGCGGCGGCCAACCAGAAAAC 165

Qy    121 AGCCCGGACCGCGTCAAGCGGCCCATGAATGCCTTCATGGTGTGGTCCCGCGGCCGCGG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    166 AGCCCGGACCGCGTCAAGCGGCCCATGAATGCCTTCATGGTGTGGTCCCGCGGCCGCGG 225

Qy    181 CGGAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    226 CGGAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC 285

Qy    241 GCCGAGTGGAAACTTTTGTCGGAGGAGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    286 GCCGAGTGGAAACTTTTGTCGGAGGAGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGG 345

Qy    301 CTGCGCGCGCTGCACATGAAGGAGCACCCGGATTATAAATACCGGCCCCG 350
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 CTGCGAGCGCTGCACATGAAGGAGCACCCGGATTATAAATACCGGCCCCG 395
```

```
RESULT 15
AF022928
LOCUS       AF022928        977 bp    mRNA       linear    VRT 20-APR-1998
DEFINITION  Xenopus laevis Sry-related HMG factor (Sox-2) mRNA, complete cds.
ACCESSION   AF022928
VERSION     AF022928.1 GI:3064135
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 977)
  AUTHORS   Mizuseki,K., Kishi,M., Matsui,M., Nakanishi,S. and Sasai,Y.
  TITLE     Xenopus Zic-related-1 and Sox-2, two factors induced by chordin,
            have distinct activities in the initiation of neural induction
  JOURNAL   Development 125 (4), 579-587 (1998)
  MEDLINE   98167885
REFERENCE   2 (bases 1 to 977)
  AUTHORS   Mizuseki,K., Matsui,M., Kishi,M., Nakanishi,S. and Sasai,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-1997) University Faculty of Medicine, Yoshida, Sakyo, Kyoto 606, Japan
            Department of Biological Sciences, Kyoto
FEATURES             Location/Qualifiers
     source          1..977
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
```

```
gene            1..977
                /gene="Sox-2"
CDS             38..973
                /gene="Sox-2"
                /note="Xenopus Sox-2"
                /codon_start=1
                /product="Sry-related HMG factor"
                /protein_id="AAC14215.1"
                /db_xref="GI:3064136"
                /translation="MYSMMETELKPPAPQQPSGGNSNSASNNQNKNSPDRVKRPMNAF
                MVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSEAEKRPFIDEAKRLRALHMKEHP
                DYKYRPRRKTKTLMKKDKYTLPGGLLAPGANAMTSGVGGSLGAGVNQRMDTYAHMNGW
                TNGYGMIEDDVGYPQHPGLNAHNAPQMLPMHRYDVSALQYNSMSSSQTYMNRSPTYS
                MSYSQQGAPGMSLGSMGSVVKSESSSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAE
                VPESAAQSRLHMSQHYQSASVAGTGINGTLPLSHM"
BASE COUNT      245 a    298 c    277 g    157 t
ORIGIN
```

Query Match   57.8%; Score 288.8; DB 5; Length 977;
Best Local Similarity 77.0%; Pred. No. 2.8e-39;
Matches 385; Conservative 0; Mismatches 97; Indels 18; Gaps 2;

```
Qy    1 CACAGCGCCCGCATGTACACAATGATGGAGACGGAGCTGAAGCCGCCGGCCCCGCAGCAA   60
          ||| |||||   |||||||| ||||||||||||  ||||| ||||| | |  |||||||
Db   26 CACGCCGCCTTCGGATGTACACCATGATGGAGACGACCGAGCTCAAGCCCCCCGCCCCGCAG   85

Qy   61 ACTTCGGGGGGGGGGGCGGGGCGGGGCCGCAACTCCACCCGCCGGCGGCCAACCAGAAAAC  120
          | |||||    ||||     ||||||||||||  |          | | |||||||| |||
Db   86 CCATCGGGGGGGGCAACTCCAACT------------CTGCGTCCAACACCAGAATAAGAAC  133

Qy  121 AGCCCGGACCGGCGTCAAGCGCCCCATGAAATGCCTTCATGGGTGTGGTCCCGGCGCAGCGG  180
          |||||||||||    ||| |||| ||||||  ||||| ||||  | |||||  |  ||  ||
Db  134 AGCCCGGACCGGAGAAGAGACCCATGAACGCCTTTCATGGTATGGTCCAGGGGCCCGGGAGG  193

Qy  181 CGCAAGAGATGGCCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGC  240
          || |||||||||| | |||||||||||||||||||||||||||||||||||||| || |||
Db  194 AGAAAGAGATGGCTCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGAGGCTGGGG  253

Qy  241 GCCGAGTGGAAACTTTTGTCGGGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGG  300
          || ||||||||  || | |  ||  || |||||     |||||||||||||||| ||| ||
Db  254 GCTGAGTGGAAGCTCCTCTCCGAGGCCGAGAAAAGACCCTTCATCGACGAGGCCAAGAGG  313

Qy  301 CTGCGAGCGCCTGCACATGAAGGAGCACCCGGATTTATAAATACCGGCCCCGCCGGAAAACC  360
          |||||||| | ||||||||||| ||| | |||||| ||||||||||||||| | ||||||||
Db  314 CTGCGAGCTCTGCACATGAAAGAGCATCCGGATTATAAATACCGGCCCCGGAGGAAAACC  373

Qy  361 AAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGCCTGCTGGCCCCCGGCGCGGC  420
          || ||  ||||||||||||||||||||||  |||||||| |  | |||||  |  |||||||
Db  374 AAAACCCTCATGAAGAAGGATAAGTACACTCTGCCCGGGGGGTTTGCTGGCTCCTGGGGGGC  433

Qy  421 AATAGCATGGCGAGCGGGGTCGGGGTGGGGCCGGGCCTGGGCGGTGAACCAGCGC  480
          || | ||||||   | |||||||||         ||  ||  | |  |||||||  |  ||
Db  434 AACGCCATGACTTCTGGGGTCGGGG--------GGCAGCGTCGGGGGCGCGGGGTCAACCAGAGG  487

Qy  481 ATGGACAGTTACGCGCCACAT  500
          ||||| ||| || ||| ||||
Db  488 ATGGACACTTATGCCCACAT  507
```

Search completed: October 10, 2002, 17:28:49
Job time : 778.867 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
(without alignments)
14616.225 Million cell updates/sec

Title: US-09-489-101A-4_COPY_1_500
Perfect score: 500
Sequence: 1 ccggccgtctatgctccagg.........ctcctggcgcggccggcgcggg 500
Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0
Searched: 1797656 seqs, 1046326829 3 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

```
 1: gb_ba:*
 2: gb_htg:*
 3: gb_in:*
 4: gb_om:*
 5: gb_ov:*
 6: gb_pat:*
 7: gb_ph:*
 8: gb_pl:*
 9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	4091	6	AX201583	AX201583 Sequence
2	500	100.0	4091	9	HSSOX1	Y13436 Homo sapien
3	500	100.0	4091	9	AL138691	AL138691 Human DNA
4	488	97.6	1542	6	AX001335	AX001335 Sequence
5 c	463.2	92.6	2376	6	AX001334	AX001334 Sequence
6	463.2	92.6	2376	10	MMSOX1	X94126 M.musculus
7	328	65.6	2312	5	AB011802	AB011802 Gallus ga
8	328	65.6	2312	6	AX001333	AX001333 Sequence
9 c	292.2	58.4	1161	6	AX001337	AX001337 Sequence
10	277	55.4	1150	10	MMSOX3	X94125 M.musculus
11	277	55.4	16632	10	AF434675	AF434675 Mus muscu
12	273.2	54.6	855	4	S69429	S69429 testis-dete
13	267.6	53.5	1824	5	AB011803	AB011803 Gallus ga
14	264.4	52.9	832	9	HSA327010	AJ327010 Homo sapi
15	264.4	52.9	1085	6	AX201582	AX201582 Sequence
16	264.4	52.9	1085	9	HSSOX2G	Z31560 H.sapiens s
17	264.4	52.9	1109	9	HUMHMGBOX	L07335 Homo sapien
18	264.4	52.9	1181	9	BC013923	BC013923 Homo sapi
19	264.4	52.9	2418	10	MMSOX2	X94127 M.musculus
20	262.8	52.6	2283	10	AX305428	AX305428 Sequence
21	262.8	52.6	2283	10	MMU31967	U31967 Mus musculu
22 c	262.8	52.6	203344	2	AL606746	AL606746 Mus muscu
23	259.6	51.9	749	9	HSA324649	AJ324649 Homo sapi
24 c	258.4	51.7	664	9	HSA322962	HSA322962 Homo sapi
25	258.4	51.7	1692	9	AF264713	AF264713 Homo sapi
26	258.4	51.7	2509	6	AX201590	AX201590 Sequence
27	258.4	51.7	2509	9	HSSOX3	X71135 H.sapiens s
28 c	258.4	51.7	148598	9	HSBA51C14	AL121875 Human DNA
29	255.8	51.2	3054	4	OASOX2GEN	X96997 O.aries SOX
30 c	246.2	49.2	1808	9	GGU12467	U12467 Gallus gall
31	230.2	46.0	180538	2	AL137061	AL137061 Human DNA
32 c	228.6	45.7	8372	6	AX201591	AX201591 Sequence
33	228.6	45.7	8372	9	AF107044	AF107044 Homo sapi
34	228	45.6	2217	5	AB026623	AB026623 Gallus ga
35	227	45.4	1186	9	CHKSOX2	D50603 Chicken mRN
36	227	45.4	1355	5	GG012532	U12532 Gallus gall
37	225.8	45.2	1725	10	AY069926	AY069926 Mus muscu
38	220.8	44.2	412	6	AX321149	AX321149 Sequence
39	220.0	44.1	395	6	AX321190	AX321190 Sequence
40	215.6	43.1	1786	5	AX245396	AX245396 Sequence
41	213.6	42.7	977	9	OIA245396	AJ245396 Oryzias 1
42	212.2	42.4	1190	5	AF022928	AF022928 Xenopus 1
43	210.4	42.1	957	9	AF005476	AF005476 Xenopus 1
44	210.4	42.1	2051	9	AF193760	AF193760 Gallus ga
45	210.4	42.1	2315	5	AB026622	AB026622 Gallus ga

ALIGNMENTS

RESULT 1
AX201583

```
LOCUS       AX201583      4091 bp   DNA    linear   PAT  30-AUG-2001
DEFINITION  Sequence 4 from Patent WO0153349.
ACCESSION   AX201583
VERSION     AX201583.1  GI:15391431
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4091)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE     Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 4 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..4091
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

RESULT 2

```
LOCUS       HSSOX1      4091 bp    DNA     linear   PRI 09-FEB-2001
DEFINITION  Homo sapiens sox1 gene.
ACCESSION   Y13436
VERSION     Y13436.1  GI:4128158
KEYWORDS    SOX1 gene; Sry-related Box 1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4091)
  AUTHORS   Malas,S., Duthie,S.M., Mohri,F., Lovell-Badge,R. and Episkopou,V.
  TITLE     Cloning and mapping of the human SOX1: a highly conserved gene
            expressed in the developing brain
  JOURNAL   Mamm. Genome 8 (11), 866-868 (1997)
  MEDLINE   98051911
REFERENCE   2  (bases 1 to 4091)
  AUTHORS   Malas,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1997) S. Malas, MRC Clinical Sciences Centre,
            Mouse Embryology, Du Cane Rd, London, W12 ONN, London, UK
REFERENCE   3  (bases 1 to 4091)
  AUTHORS   Malas,S.
  REMARK    Revised by [3]
  JOURNAL   Submitted (06-JAN-1999) S. Malas, MRC Clinical Sciences Centre,
            Mouse Embryology, Du Cane Rd, London, W12 ONN, London, UK
COMMENT     On Jan 8, 1999 this sequence version replaced gi:2230882.
FEATURES             Location/Qualifiers
     source          1..4091
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="q33-34"
                     /clone="psxBg1.1"
     gene            61..1224
                     /gene="SOX1"
     CDS             61..1224
                     /gene="SOX1"
                     /codon_start=1
                     /product="sry-related Box 1 protein"
                     /protein_id="CAA73847.1"
                     /db_xref="GI:2230883"
                     /db_xref="SWISS-PROT:000570"
                     /translation="MYSMMMETDLHSPGGAQAPTNLSGPAGAGGGGGGGGGGGGA
                     KANQDRVKRPMNAFMVWSRGQRRKMAQENPRKMHNSEISKRLGAEWKMMSEAEKRPFI
                     EAKRLRALHMKEHPDYKYRPRRKTKTLLKKDKYSLAGGLLAAGAGGGGAAVAMGVGVG
                     VGAAPVGQRLESPGGAAGGYAHVNGWANGAYPGSVAAAAAAAAMMQEAQLAYGQHPG
                     AGGAHPHRTPAHPHPHPHAHPHNPQPMHRYDMGALQYSPISNSQGYMSASPSGYGGL
                     PYGAAAAAAAAHQNSAVAAAAAAASSGALGALGSLVKSEPSGSSPPAPAHSRAPCPG
                     DLREMISMYLPAGEGDPAAAAAAAAAQSRLHSLPQHYQGAGAVNGTVPLIHI"

Related sequences: AI279621, AI298071, AI215744, AA960996, D81624,
R24723, R20579, T07302, R14439, AA961095, T06325, R46080.

BASE COUNT      924 a    1109 c    1110 g     943 t      5 others
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 4091;
  Best Local Similarity  100.0%;  Pred. No. 6.5e-51;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGGCCGTCTATGCTCCAGGCCCTCTCCTCTCGCGGTGCCGGTGCCGGTGCCCGCCCCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCGGCCGTCTATGCTCCAGGCCCTCTCCTCTCGCGGTGCCGGTGCCGGTGCCCGCCCCG 60

Qy      61 ATGTACAGCATGATGATGATGGAGACCGACCTGCACTCGCCCGGCGGGGCCCCACG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATGTACAGCATGATGATGATGGAGACCGACCTGCACTCGCCCGGCGGGGCCCCACG 120

Qy     121 AACCTCTCGGGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AACCTCTCGGGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC 180

Qy     181 GGCGGGGCCGGCCAACCAGGACCGGGGGTCAAACGGCCCATGAACGCCTTCATGGTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCGGGGCCGGCCAACCAGGACCGGGGGTCAAACGGCCCATGAACGCCTTCATGGTG 240

Qy     241 TGGTCCCGGGGCCAGAGGAGGAAGATGGCCCAGGAGAACCCCAGAAACGCCCATAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGTCCCGGGGCCAGAGGAGGAAGATGGCCCAGGAGAACCCCAGAAACGCCCATAG 300

Qy     301 ATCAGCAACGCGCTGGGGGCCGAGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATCAGCAACGCGCTGGGGGCCGAGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCC 360

Qy     361 ATCGACGAGGCCCAAGCGGCTGCGCGCCCTGCACATGAAGGAGCACCCGGATTACAAGTAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATCGACGAGGCCCAAGCGGCTGCGCGCCCTGCACATGAAGGAGCACCCGGATTACAAGTAC 420

Qy     421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGGCCGGGGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGGCCGGGGG 480

Qy     481 CTCCTGGCGGCGGCGGCGGGG 500
           |||||||||||||||||||||
Db     481 CTCCTGGCGGCGGCGGCGGGG 500
```

RESULT 3

```
LOCUS       AL138691
DEFINITION  Human DNA sequence from clone RP11-310D8 on chromosome 13, complete
LOCUS       AL138691/c  147665 bp   DNA     linear   PRI 16-JUN-2001

BASE COUNT      924 a    1109 c    1110 g     943 t      5 others
ORIGIN

Query Match             100.0%;  Score 500;  DB 9;  Length 4091;
  Best Local Similarity  100.0%;  Pred. No. 6.5e-51;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGGCCGTCTATGCTCCAGGCCCTCTCCTCTCGCGGTGCCGGTGCCGGTGCCCGCCCCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCGGCCGTCTATGCTCCAGGCCCTCTCCTCTCGCGGTGCCGGTGCCGGTGCCCGCCCCG 60

Qy      61 ATGTACAGCATGATGATGATGGAGACCGACCTGCACTCGCCCGGCGGGGCCCCACG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATGTACAGCATGATGATGATGGAGACCGACCTGCACTCGCCCGGCGGGGCCCCACG 120

Qy     121 AACCTCTCGGGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AACCTCTCGGGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC 180

Qy     181 GGCGGGGCCGGCCAACCAGGACCGGGGGTCAAACGGCCCATGAACGCCTTCATGGTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGCGGGGCCGGCCAACCAGGACCGGGGGTCAAACGGCCCATGAACGCCTTCATGGTG 240

Qy     241 TGGTCCCGGGGCCAGAGGAGGAAGATGGCCCAGGAGAACCCCAGAAACGCCCATAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGTCCCGGGGCCAGAGGAGGAAGATGGCCCAGGAGAACCCCAGAAACGCCCATAG 300

Qy     301 ATCAGCAACGCGCTGGGGGCCGAGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATCAGCAACGCGCTGGGGGCCGAGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCC 360

Qy     361 ATCGACGAGGCCCAAGCGGCTGCGCGCCCTGCACATGAAGGAGCACCCGGATTACAAGTAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATCGACGAGGCCCAAGCGGCTGCGCGCCCTGCACATGAAGGAGCACCCGGATTACAAGTAC 420

Qy     421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGGCCGGGGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCTGGCCGGGGG 480

Qy     481 CTCCTGGCGGCGGCGGCGGGG 500
           |||||||||||||||||||||
Db     481 CTCCTGGCGGCGGCGGCGGGG 500
```

sequence.
ACCESSION AL138691
VERSION   AL138691.15 GI:14485290
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 147665)
  AUTHORS   Tromans,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequest@sanger.ac.uk
COMMENT   On Jun 17, 2001 this sequence version replaced gi:13184286.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. The following
          abbreviations are used to associate primary accession numbers given
          in the feature table with their source databases: Em:, EMBL; Sw:,
          SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome 13. Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr13
          Mapping Group.
          RP11-310D8 is from the library RPCI-11.2 constructed by the group
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          This sequence is the entire insert of clone RP11-310D8 The true
          left end of clone RP11-75F3 is at 53050 in this sequence. The true
          right end of clone RP11-450H6 is at 21465 in this sequence.
FEATURES   Location/Qualifiers
    source   1..147665
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="13"
              /clone="RP11-310D8"
              /clone_lib="RPCI-11.2"
    misc_feature   4735
              /note="Tandem repeat. Forced join. Assembly confirmed by
              restriction digest data."
    misc_feature   4736..4759
              /note="Single clone region. Sequence from reads from a
              short insert library derived from a clone PCR.
              Restriction digest data confirm the assembly."
    misc_feature   4760..5371
              /note="Sequence from clone PCR only."
    misc_feature   14935..15023
              /note="Single clone region. Assembly confirmed by
              restriction digest data."
    unsure   15017
BASE COUNT   38628 a   35708 c   35003 g   38326 t
ORIGIN

Query Match   100.0%; Score 500; DB 9; Length 147665;
Best Local Similarity 100.0%; Pred. No. 2.7e-51;
  Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CCGGCGCGTCTATGCTCCAGGCCCTCTTCCTCGCGGTTGAACCCGCCAGCGCCCACG 60
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 15907 CCGGCGCGTCTATGCTCCAGGCCCTCTTCCTCGCGGTTGAACCCGCCAGCGCCCACG 15848
Qy   61 ATGTACACGCATGATGATGGAGAGACCTTGCACTTCGCCCGGGGCGGCCCCACG 120
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15847 ATGTACACGCATGATGATGGAGAGACCTTGCACTTCGCCCGGGGCGGCCCCACG 15788
Qy   121 AACCTTCTCGGGCCCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15787 AACCTCTCGGGCCCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 15728
Qy   181 GCGGGGGGGGGCGCCAAGGCCAACCAGGAACCGGGGTCAAACGGCCTTCATGGTG 240
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15727 GCGGGGGGGGGCGCCAAGGCCAACCAGGAACCGGGGTCAAACGGCCTTCATGGTG 15668
Qy   241 TGGTCCCCGCGGGCAGCGGCGGCCAAGATGGCCCAGGAGAACCCCAGAGTGCAC 300
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15667 TGGTCCCCGCGGGCAGCGGCGGCCAAGATGGCCCAGGAGAACCCCAGAGTGCAC 15608
Qy   301 ATCAGCAAGCACGCGCCTTGGGGGCCGAGTGGGAAGGTCATGTCCAGGGCGGAG 360
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15607 ATCAGCAAGCACGCGCCTTGGGGGCCGAGTGGGAAGGTCATGTCCAGGGCGGAG 15548
Qy   361 ATCGGACGAGGGCCAAGCGGCGTGCCGGCCGCTTCCACATGACCCGGATTACAAGTAC 420
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15547 ATCGGACGAGGGCCAAGCGGCGTGCCGGCCGCTTCCACATGACCCGGATTACAAGTAC 15488
Qy   421 CGGCGCGGCGCGCCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCGTGGCGGGG 480
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 15487 CGGCGCGGCGCGCCAAGACCAAGACGCTGCTCAAGAAGGACAAGTACTCGCGTGGCGGGG 15428
Qy   481 CTCCTGGCGCGGCGGCGGCGGGG 500
      |||||||||||||||||||||||
Db 15427 CTCCTGGCGCGGCGGCGGCGGGG 15408

RESULT 4
LOCUS   AX001335   1542 bp   DNA   linear   PAT 10-MAR-2000
DEFINITION Sequence 3 from Patent WO9900516.
ACCESSION AX001335
VERSION   AX001335.1 GI:7241519
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1542)
  AUTHORS   Pevny,L.H. and Smith,A.
  TITLE   NEURONAL STEM CELL GENE
  JOURNAL   Patent: WO 9900516-A 3 07-JAN-1999;
          MEDICAL RES COUNCIL (GB); PEVNY LARYSSA H (GB)
FEATURES   Location/Qualifiers
    source   1..1542
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
    CDS   60..1223
              /note="unnamed protein product"
              /codon_start=1
              /protein_id="CAB77114.1"
              /db_xref="GI:7241520"
              /translation="MYSMMMETDLHSPGGAQAPTNLSPAGSAGGGGGGGGGGGGGGGGGGA
              KANQDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKRPFID
              EAKRLRALHMKEHPDYKYRPRRKTKTLLKKDKYSLAGGLLAAGAGGAAVAMGVGV
              VGAAPVGQRLESPGGAAGGAYAHVNGWANGAYPGSVAAAAAAAAAMMQEAQLAYGQHPG
              PYGAAAAAAAAAHQNSAVAAAAAAASSGALGALGSLVKSEPSGSPPAPAHSRAPCPG
              AGGAHPRTPAHPHPHIPHAHPHNPQPMHRYDMGALQYSPISNSQGYMSASPSGYGL
              DLREMISMYLPAGEGGDPAAAAAAAAQSRLHSLPQHYQGAGAGVNGTVPLTHI"
BASE COUNT   238 a   565 c   560 g   179 t
ORIGIN

Query Match   97.6%; Score 488; DB 6; Length 1542;
Best Local Similarity 99.8%; Pred. No. 2.2e-49;
  Matches 499; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

RESULT 5

```
LOCUS       AX001334    2376 bp    DNA    linear   PAT 10-MAR-2000
DEFINITION  Sequence 2 from Patent WO9900516.
ACCESSION   AX001334
VERSION     AX001334.1  GI:7241518
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2376)
  AUTHORS   Pevny,L.H. and Smith,A.
  TITLE     NEURONAL STEM CELL GENE.
  JOURNAL   Patent: WO 9900516-A 2 07-JAN-1999;
            MEDICAL RES COUNCIL (GB); PEVNY LARYSSA H (GB)
FEATURES             Location/Qualifiers
     source          1..2376
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT      439 a    772 c    745 g    420 t
ORIGIN
```

Query Match   92.6%; Score 463.2; DB 6; Length 2376;
Best Local Similarity 95.4%; Pred. No. 1.7e-46;
Matches 477; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
Qy      1  CCGGCCGTCTATGCTCCAGGCCCTCTTCCTCCGGTGAAACCCGCCACGGCCCG        60
           |||||||||||| ||||||||||||||||||| ||||| |||||||||| || ||
Db      1  CCGGCCGTCTAT-CTCCAGGCCCTCTTCCTCGCGGTGCCGGTGAAACCCCGCCAGCGCCCG  59

Qy     61  ATGTACAGCATGATGATGAGGAGGACCGACCTGCACTCGCCGCCACGGCCCCACG       120
           ||||||||||||||| |||||||||||||||||||||||||| |||||||| ||||
Db     60  ATGTACAGCATGATGAGGAGGACCGACCTGCACTCGCCGCCGCCCCAGCCCCACG      119

Qy    121  AACCTCTCGGGCCCCGCCGGGGCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC    180
           |||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db    120  AACCTCTCGGGCCCCGCCGGGGCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC    179

Qy    181  GGCGGGGGGGGCCCAAGGCCAACCAGGACCGGGGGGGGGGGGGGGGGGGGGGGGGG    240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180  GGCGGGGGGGGCCCAAGGCCAACCAGGACCGGGGGGGGGGGGGGGGGGGGGGGGGG    239

Qy    241  TGGTCCCGGCAGCGGGGCGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAG   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240  TGGTCCCGGCAGCGGGGCGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAG   299

Qy    301  ATCAGCAAGCGCCTGGGGGCCGAGTGGAAGGTTCATGTCCGAGGCGGGCCGGTTC       360
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    300  ATCAGCAAGCGCCTGGGGGCCGAGTGGAAGGTCATGTCCGAGGCGGGCCGGTTC        359

Qy    361  ATCGACGAGGCCAAGCGGCCGTTCGGGGGCGGTGGAAGGTCATGTCCGAGAAGAGCGCGGTTC   420
           ||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db    360  ATCGACGAGGCCAAGCGGCTGCGCGCCCTGCACATGAACGCCTTCATGGTG          419

Qy    421  CGGCGCCGCCGCCGAAGACCAAGAAGCTGCTCTCAAGAAGAGACTACTCGGGCCGGG      480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420  CGGCGCCGCCGCCGAAGACCAAGAAGCTGCTCTCAAGAAGAGACTACTCGGGCCGGG     479

Qy    481  CTCCTGGCGGCGGCGGCGGG        500
           ||||||||||||||||||||
Db    480  CTCCTGGCGGCGGCGGCGGG        499
```

RESULT 6

```
LOCUS       MMSOX1      2376 bp    DNA    linear   ROD 13-MAR-1996
DEFINITION  M.musculus SOX1 gene.
ACCESSION   X94126
VERSION     X94126.1  GI:1213024
KEYWORDS    sox-1 gene; SOX1 protein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2376)
  AUTHORS   Collignon,J., Sockanathan,S., Hacker,A., Cohen-Tannoudji,M.,
            Norris,D., Rastan,S., Stevanovic,M., Goodfellow,P.N. and
            Lovell-Badge,R.
  TITLE     A comparison of the properties of sox-3 with Sry and two related
            genes, Sox-1 and Sox-2
  JOURNAL   Development 122 (2), 509-520 (1996)
  MEDLINE   96189340
REFERENCE   2  (bases 1 to 2376)
  AUTHORS   Sockanathan,T.E.L.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1995) TEL.S. Sockanathan, National Institute For
            Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
COMMENT     Overlaps with X55491.
FEATURES             Location/Qualifiers
     source          1..2376
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /dev_stage="8.5 dpc"
     gene            436..1611
                     /gene="sox-1"
     CDS             436..1611
                     /gene="sox-1"
                     /codon_start=1
                     /function="transcription factor"
                     /product="SOX1 protein"
                     /protein_id="CAA63846.1"
                     /db_xref="GI:1213025"
                     /db_xref="MGD:MGI:98357"
                     /db_xref="SWISS-PROT:P53783"
                     /translation="MYSMMMETDLHSPGGAQAPTNLSGPAGARGGGVGGGGGGG
                     KANQDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKRPFID"
```

```
Qy    121  AACCTCTCGGGCCCCGCCGGGGCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC    180
           |||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db    496  AACCTCTCGGGCCCCGCCGGGGCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGC    555

Qy    181  GGCGGGGGGGGCCCAAGGCCAACCAGGACCGGGGGGGGGTGGGGGCGGC             240
           |||||||||||||||||||||||||||||||||||||
Db    556  GGCGGGGGGGGCCCAAGGCCAACCAGGACCGGGGGGGGGTGGGGGCGGC             615

Qy    241  TGGTCCCGGCAGCGGGGCGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAG   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    616  TGGTCCCGGCAGCGGGGCGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAG   675

Qy    301  ATCAGCAAGCGCCTGGGGGCCGAGTGGAAGGTCATGTCCGAGAAAACCCCAAGATGCACAACTCGGAG   360
           ||||||||||||||||||||||||||||||
Db    676  ATCAGCAAGCGCCTGGGGGCCGAGTGGAAGGTCATGTCCGAG                    735

Qy    361  ATCGACGAGGCCAAGCGGCTGCTGCTCTCAAGAAGAGACTACTCGGGCCGGG          420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    736  ATCGACGAGGCCAAGCGGCTGCTGCTCTCAAGAAGAGACTACTCGGGCCGGG          795

Qy    421  CGGCGCCGCCGCCGAAGACCAAGAAGCTGCTCTCAAGAAGAGACTACTCGGGCCGGG      480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    796  CGGCGCCGCCGCCGAAGACCAAGAAGCTGCTCTCAAGAAGAGACTACTCGGGCCGGG     855

Qy    481  CTCCTGGCGGCGGCGGCGGG        500
           ||||||||||||||||||||
Db    856  CTGCTAGCGGCGGCGGCGGG        875
```

```
misc_feature   579..816
               /gene="sox-1"
               /note="HMG box"
```

```
EAKRLRALHMKEHPDYKYRPRRKTKTLLKKDKYSLAGGLLAAGAGGGGGAAVAMGVGVG
VGAAAVGQRLESPGGAAGGGYAHVNGWANGAYPGSVAAAAAAAAMMQEAQLAYGQHPG
AGGRIPHAHPHAHPHPHPHPHAHPHNPQPMHRYDMGALQYSPISNSOGYMSASPSGYGGI
PYGAAAAAAAAGGAHQNSAVAAAAAAAAASSGALGALGSLVKSEPSSPPAHSRA
PCPGDLREMISMYLPAGEGGDPAAAAAAAAQSRLHSLPQHYQGAGAVGNGTVPLTHI"
```

```
BASE COUNT     439 a    772 c    745 g    420 t
ORIGIN

Query Match    92.6%;   Score 463.2;   DB 10;   Length 2376;
Best Local Similarity   95.4%;   Pred. No. 1.7e-46;
Matches  477;   Conservative  0;   Mismatches  23;   Indels  0;   Gaps  0;

Qy      1 CCGGCCGTCTATGCTCCAGGCCCCTCTCTCCTCGCGGCTGCCGGTGACCGGCCAGCCCCG   60
          ||||||||| |||||||||||| |||| |||||||| | ||| |||||||||| |||||
Db    376 CCGGGCGTTCTATGCTCCAGGCCCCTCTTCGCCGTACCGGTGAACCGGCCAGCCCCAG   435

Qy     61 ATGTACAGCATGATGATGGAGAGACCGACCGACCGCCTCCTTGCGGTGGTGGTGGTGGKAG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 ATGTACAGCATGATGATGGAGAGACCGACCGACCGCCTCCTTGCGGTGGTGGTGGTGGCACG   495

Qy    121 AACCTCTCGGGCCCCGGGGGCCCCGGCGGGGGCGGGGGGGGCGGGGGCGGGGGCGGGGGGC  180
          |||||||||||||||||||||||
Db    496 AACCTCTCGGGCCCCGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGC   555

Qy    181 GGGGGGGGGCGCCAAGGCCAACCAGGATCGGGTCAAACGGCCCATGAACGCCTTCATGGTG  240
          ||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||
Db    556 GGGGGGGGCCACCAAGGCCAACCAGGATCGGGTCAAGCGGCCCATGAACGCCTTCATGGTG  615

Qy    241 TGGTCCCGCGGCCAGCGCCGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAG  300
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    616 TGGTCCCGCGGCCAGCGGCGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAG  675

Qy    301 ATCAGCAAGCGCCTGGGGGCGGAGTGGAAGGTCATGTCGGAGGCGGAGAAGCGGCCGTTC  360
          |||||||||||||| |||||| |||||||||||||||| |||||| ||||||||||||||
Db    676 ATCAGCAAGCGCCTCGGGGCCGAGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTC  735

Qy    361 ATCGACGAGGCGAAGCGGCTGCGCGCGCTGCACATGAAGGAGCACCCCGGATTACAAGTAC  420
          |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db    736 ATCGACGAGGCGAAGCGGCTGCGCGCGCTGCACATGAAGGAACACCCCGATTACAAGTAC  795

Qy    421 CGGCCGCGCCGCAAGACCAAGAAAGCACGGTTGCTCAAGAAGGACAAGTACTCGCTGGCCG  480
          |||||||||||||||||||||  |||   |||  ||||||||||||||||||||||||||
Db    796 CGGCCGCGCCGCAAGACCAAGACTGCGCGCGCTGCACATGAAGGACAAGTACTCGCTGGCCG   855

Qy    481 CTCCTGGCGGCGCCGCGGGGG   500
          || |||||||||||||||||||
Db    856 CTGCTGGCGGCGCCGCGGGGG   875
```

```
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /gene="Sox1"
                     /dev_stage="embryo"
     gene            209..1330
                     /gene="Sox1"
     CDS             209..1330
                     /gene="Sox1"
                     /function="SRY-type HMG domain containing transcription
                     factor"
                     /codon_start=1
                     /product="SOX1"
                     /protein_id="BAA25092.1"
                     /db_xref="GI:2947025"
                     /translation="MYSMMMETDLHSPGGAPAPGGGLSGQSGGAGGGGGGGKAG
                     QDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKVMSEAEKRPFIDEAK
                     RLRALHMKEHPDYKYRPRRKTKTLLKKDKYSLAGGLLAGGLLAGPAAGGPPAVGVMGVGVI
                     PGGVGQRLESPGGAAGGGYAHMNGWANGAYPGSVAAAAAAAMMQEAQLAYGQPHQ
                     GHPHPHPHPHPHPHPHRYDMGALQYSPISNSQGYVSASPSGYGALPYGSAQP
                     NSAAAAAAAAAASSGALGALGSLVKSEPSSPPVTSHRAPCPGDLREMISMYLP
                     GGEGGDPAAAAAQSRLHSLPQHYQSASTGVNGTVPLTHI"
```

```
BASE COUNT     421 a    739 c    714 g    438 t
ORIGIN

Query Match    65.6%;   Score 328;   DB 5;   Length 2312;
Best Local Similarity   81.6%;   Pred. No. 1.6e-30;
Matches  395;   Conservative  0;   Mismatches  80;   Indels  9;   Gaps  1;

Qy     16 CCAGGCCCTCTCCTCGCGGCTGCCGGTGACCGGCCAGCCCCG   75
          || |||| ||||||||||| |||  | ||  |||||||||| ||||
Db    164 CCTCCCGCGCCCCGCCCGCCGGGAGGCAAGTGAGGAAGTGACGAAGTACAGCAGCATGATG   223

Qy     76 ATGGAGAGACCGACTTGCACTCGCCCCGGCCCCCCAGCCCC------CCGGGCGGCGGC  135
          |||| | || |||  ||||||||||||||||||||||||||      ||||||||| |
Db    224 ATGGACAACGGACTTGCACTCGCCCCCGGCCCCCCAGCCGAGAGCCCCGGGGCGGCGGC   274

Qy    136 GCCGGGGGGCGGCGGGGGGCGGGGGGGGGGGGGGGGGCCAAG   195
          ||| || |||| |||| || ||||||||| |||||||||||
Db    275 CTCTCGGGGCGGCAGAGCGGCGGGCGGGGGGGGGGGGGGCAAA   334

Qy    196 GCCAACCAGGACGGGGTCAAACGGCCCATGAACGCCTTCATGGTGTGGTCCCGCGGCCAG   255
          || |||||||| ||| ||||  |||||||| |||||||||||||||||||||||||||||
Db    335 GCGGGGCAGGACCGGGTGAAGCGCCCCATGAACGCCTTCATGGTGTGGTCGCGGGGGCAG   394

Qy    256 CGGCGCAAGATGGCCCAGGAGAATCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTG   315
          ||||| |||||| || || ||||  |||||||||||| ||||| |||||||||||| ||
Db    395 CGGCGGAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGGCTG   454

Qy    316 GGGGCCGAGTGGAAGGTCATGTCGGAGGCCGAGAAGCGGCCGTTCATCGACGAGGCCAAG   375
          ||||||||||||||||||||||| ||||||||||||||||||||||||||| ||| |||
Db    455 GGCGCCGAGTGGAAGGTGATGTCGGAGGCCGAGAAGCGGCCCTTTCATCGACGAGGCGAAG   514

Qy    376 CGGCTGCGCGCGCTGCACATGAAGGAGCACCCGGATTACAAGTACCGGCCGCGCCGCAAG   435
          |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||| |||
Db    515 CGGCTGCGCGCGCTGCACATGAAGGAGCACCCGGATTATAAATACCGGCCGCGCCGGAAG   574

Qy    436 ACCAAGACGCTGCTCAAGAAGGACACAAAGTACTCGCTGGCCGGCGGCCTGCTGGCCGGC   495
          ||||||||| |||||||||||||||| || |||||||||||||| | || ||| ||||
Db    575 ACCAAGACCCTGCTCAAGAAGGACAAGTACTCGCTGGCCGGCGGCCTGCTGGGCGGCGGC   634

Qy    496 GCGG   499
          ||||
Db    635 CCGG   638
```

```
LOCUS       AX001333          2312 bp    DNA    linear   PAT 10-MAR-2000
DEFINITION  Sequence 1 from Patent WO900516.
ACCESSION   AX001333
VERSION     AX001333.1  GI:7241517
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2312)
  AUTHORS   Pevny,L.H. and Smith,A.
  TITLE     NEURONAL STEM CELL GENE
  JOURNAL   Patent: WO 9900516-A 1 07-JAN-1999;
            MEDICAL RES COUNCIL (GB); PEVNY LARYSSA H (GB)
FEATURES             Location/Qualifiers
     source          1..2312
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
BASE COUNT      421 a    739 c    714 g    438 t
ORIGIN
```

```
Query Match  65.6%;  Score 328;  DB 6;  Length 2312;
Best Local Similarity  81.6%;  Pred. No. 1.6e-30;
Matches  395;  Conservative  0;  Mismatches  80;  Indels  9;  Gaps  1;

Qy       16  CCAGGCCCTCTCTTCGCGGTCGCGGTGAACCCGCCGAGTGTACAGCATGATG   75
             || ||   || ||| | |  ||  |  | |  |||    |  | ||||||||||
Db      164  CCTCCCCGCGCCGCGGGCGCCGGAGGGAAGTGAGGAAGCCCGTGAATGTACAGCATGATG  223

Qy       76  ATGGAGACGACCTGCACTCGCCCCGCGGGCGGGCCCAGCCCCACCTCTCGGGCCCC  135
             |||| |||||| ||  ||| ||||||||||||||   | | ||  |||
Db      224  ATGGAGACGGACTTGCACTCGCCCCGCGGGGACCCCGGGGCC--------CGGCGGCGGC  274

Qy      136  GCCGGGGCGGCGGGGGGCGCGCGGGGGGCGGGAGGCGGGGGGGGCGGGGGGGCCAAG  195
             || |||   ||| ||   | |||||  ||||||||||||||||||||| | |||
Db      275  CTCTCGGGGCCAGAGCGGCGGCGCGGGGGGCGGGGGGGGCGGCGGGGGGGCAAA  334

Qy      196  GCCAACCAGGACGCGGGGTCAAAACGGGCCCATGAACACGGCCTTCATGTGTTGGTCCCGCGGCAG  255
             ||    |||||  ||||| | | ||||||||||||||||||||| | ||
Db      335  GCGGGGCAGGACCCGCGTGAAAGCCGCCCCATGAACGCCCTTCATGGTGTCGGCGGCGGGGGCAG  394

Qy      256  CGGCGCAAGATGGCCCAGGAGAGAACCCCAAGAGATCACAACTCGGAGATCGGAGCGGCTG  315
             |||| |||||  | ||||||||||||||||||||||||||||||||| |||||||||||||
Db      395  CGGCGGGAAGATGGCCCAGGAGAATCCCAAGAGATGCACAACTCGGAGATCAGCAAGCGGCGGTG  454

Qy      316  GGGGGCCCGAGTGGAAAGGTCATGTCCGAGCCCGAGAAGCGGCCGCTTTCATCGACGAGGCCAAG  375
             || ||   ||| |||| |||| || ||||||||||||||||||||| ||
Db      455  GCGCCGAGTGGAAGGTGATGTCGGAGCCCGACGAAGCGCCGCCTTTCATCGACGAGGCGAAG  514

Qy      376  CGGCTGCGCGCGGCCTGGCACATGAAGGAGGACCACCCGGATTACAAGTACCGGCCGCGCAAG  435
             |||||||||| | | ||| |||| ||||||||| ||| ||  |||   |||   || || |||
Db      515  CGGCTGCGGGCGGCCTGCCACAACCCGGATTATAAAATACCGGCCCCCGGCGGAAG  574

Qy      436  ACCAAGAGACGCTGCTCAAGAAGGACAAGTACTCGGTCGGCTCCTCGGCCGCGGCCGGC  495
             |||||| | || |||| |||||||||||||| |||||| ||||||||||||||||||
Db      575  ACCAAGACGCTCTCAAGAAGGACAAGTACTCGCTGGGCGGGCTCCTGGGCCGCCCGC  634

Qy      496  GCGG  499
             |||
Db      635  CCGG  638
```

```
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Pevny,L.H. and Smith,A.
  TITLE     NEURONAL STEM CELL GENE
  JOURNAL   Patent: WO 9900516-A 5 07-JAN-1999;
            MEDICAL RES COUNCIL (GB); PEVNY LARYSSA H (GB)
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     variation       7..9
                     /note="AGY"
     variation       28..30
                     /note="CUN"
     variation       34..36
                     /note="AGY"
     variation       64..66
                     /note="CUN"
     variation       67..69
                     /note="CUN"
     variation       148..150
                     /note="AGY"
     variation       157..159
                     /note="CGN"
     variation       184..186
                     /note="AGY"
     variation       187..189
                     /note="AGY"
     variation       196..198
                     /note="CGN"
     variation       199..201
                     /note="CGN"
     variation       235..237
                     /note="AGY"
     variation       244..246
                     /note="AGY"
     variation       250..252
                     /note="CGN"
     variation       253..255
                     /note="CUN"
     variation       277..279
                     /note="AGY"
     variation       291..294
                     /note="CGN"
     variation       316..318
                     /note="CGN"
     variation       319..321
                     /note="CUN"
     variation       322..324
                     /note="CGN"
     variation       328..330
                     /note="CUN"
     variation       361..363
                     /note="CGN"
     variation       367..369
                     /note="CGN"
     variation       370..372
                     /note="CGN"
     variation       385..387
                     /note="CUN"
     variation       388..390
                     /note="AGY"
     variation       406..408
                     /note="CUN"
     variation       409..411
                     /note="AGY"
     variation       421..423
                     /note="CUN"
     variation       424..426
                     /note="CUN"
     variation       505..507
                     /note="CGN"
     variation       508..510
                     /note="CUN"
```

```
RESULT 9
LOCUS       AX001337          1161 bp    DNA    linear   PAT 10-MAR-2000
DEFINITION  Sequence 5 from Patent WO900516.
ACCESSION   AX001337
VERSION     AX001337.1  GI:7241521
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
```

```
variation    513..515
             /note="AGY"
variation    583..585
             /note="AGY"
variation    631..632
             /note="AGY"
variation    676..678
             /note="CUN"
variation    742..744
             /note="CGN"
variation    760..762
             /note="CGN"
variation    769..771
             /note="CUN"
variation    778..780
             /note="AGY"
variation    784..786
             /note="AGY"
variation    799..801
             /note="AGY"
variation    805..807
             /note="AGY"
variation    811..813
             /note="AGY"
variation    836..838
             /note="CUN"
variation    871..873
             /note="CUN"
variation    907..909
             /note="AGY"
variation    910..912
             /note="AGY"
variation    919..921
             /note="CUN"
variation    928..930
             /note="CUN"
variation    934..936
             /note="AGY"
variation    937..939
             /note="CUN"
variation    946..948
             /note="CUN"
variation    955..957
             /note="AGY"
variation    961..963
             /note="AGY"
variation    982..984
             /note="AGY"
variation    985..987
             /note="AGY"
variation    1006..1008
             /note="CGN"
variation    1009..1011
             /note="CUN"
variation    1021..1023
             /note="AGY"
variation    1030..1032
             /note="CUN"
variation    1084..1086
             /note="CGN"
variation    1087..1089
             /note="CGN"
variation    1090..1092
             /note="CUN"
variation    1096..1098
             /note="CUN"
variation    1099..1101
             /note="AGY"
variation    1150..1152
             /note="CUN"
BASE COUNT    173 a    216 c    276 g    130 t    366 others
ORIGIN
```

```
RESULT 10
MMSOX3
LOCUS       MMSOX3      1150 bp    DNA    linear    ROD  13-MAR-1996
DEFINITION  M.musculus SOX3 gene.
ACCESSION   X94125
VERSION     X94125.1  GI:1209431
KEYWORDS    sox-3 gene; SOX3 protein.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1150)
AUTHORS     Collignon,J., Sockanathan,S., Hacker,A., Cohen-Tannoudji,M.,
            Norris,D., Rastan,S., Stevanovic,M., Goodfellow,P.N. and
            Lovell-Badge,R.
TITLE       A comparison of the properties of Sox-3 with Sry and two related
            genes, Sox-1 and Sox-2
JOURNAL     Development 122 (2), 509-520 (1996)
MEDLINE     96189340
REFERENCE   2  (bases 1 to 1150)
AUTHORS     Sockanathan,T.E.L.S.
TITLE       Direct Submission
JOURNAL     Submitted (07-DEC-1995) TEL.S. Sockanathan, National Institute For
            Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
COMMENT     Overlaps with X55491.
FEATURES             Location/Qualifiers
     source          1..1150
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /dev_stage="8.5 dpc"
     gene            23..1150
                     /gene="sox-3"
     CDS             23..1150
                     /gene="sox-3"
                     /function="transcription factor"
```

```
Query Match    58.4%;  Score 292.2;  DB 6;  Length 1161;
Best Local Similarity 63.2%;  Pred. No. 3.3e-26;
Matches 278;  Conservative 62;  Mismatches 100;  Indels 0;  Gaps 0;

Qy   61 ATGTAACAGCATGATGATGGAGGACCGGACCTGCCACTCGCCCGGCGGCCCCCACG 120
         ||||::|  ||||||||||||||: |||  |||||||||||||||:| | || ||:
Db    1 ATGTRAYTCNATGATGATGGARACNGAYTTRCAYTCNCCNGGNGGNCNCARGCNCCNCACN 60

Qy  121 AACCTCTCGGGCCCCCGCCCGGCGGGGCGGGGGGGGGGGGGAGGCGGGGGGGGGGGGC 180
         ||||: |: || ||:||:|||||||:|||||:||||||||||:|||||||||||||
Db   61 AAYTTRTCNGGNCCNCGNCCNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGN 120

Qy  181 GGCGGGGGGGCCCAAGGCCAACCAGGACCGGGTCAAACGGCCATTCAAACGCCTTCATGGTG 240
         || || ||| || ||||| || ||| |:||| |||||:| |||:||||: |  ||||:|||
Db  121 GGNGGNGGNGCNAARGCNAAYCARGAYAGRGTNAARAGRCCNATNGAAYGCNTTYATGGTN 180

Qy  241 TGGTCCCCCGGGGGGACGGCGGGGCGGGCCAAGACAGAGGAACAACGTCACAACTCGGAG 300
         ||||| | |||||:|| || ||| |||  ||||||||| ||:||:|| || | || ||:
Db  181 TGGTCNAGRGGNCNARGAYAGRAAAAYGGCNARGARAAYCCNAARAYGCAYAAYTCNGAR 240

Qy  301 ATCAGCAACGACCTGGGGCCGACTGGGAAGGTCAGTCCGAGGCCGGAGAAGCGCGGCGTTC 360
         || |: || | |:| ||: |:||||| || || |: ||:|||||| || ||:|| |::
Db  241 ATHCNAARAGRYTRGGNCNGARTCGAARGTNAYNGTCNGARGCNGARAARAGRGCCNTTY 300

Qy  361 ATCGACGAGGCCAAGCGCCTGCCGCGGCCTTGCACATGAGGAGCACCCGGATTACAAGTAC 420
         || || || || ||||| || ||:|| |:|||:| || ||:|| ||:||| ||:||:||
Db  301 ATHGAYGARGCNAARAGRYTRAGRCNTTRCAYATGARRARCACNCGNAYNCCNAYTAYAARTAY 360

Qy  421 CGGCCGCGCCGCAAGACCAAGACGCTGCTCAAGAAGGACCAAGTACTCGCTGGCCGCGGGG 480
         ||:|| || || || ||||| |||:| || || || |||||:| || |:|||:|| |||
Db  361 AGRCCNAGRAGRAAACNAARACNTTRTTRAARAARGAYAARTAYTCNTTRGCNGNGNGGN 420

Qy  481 CTCCTGGCCGCCGGCGGCGGG 500
         |:|| || || || |||:||
Db  421 TTRTTRGCNGCNGGNGGNGNG 440
```

```
                        JOURNAL    Submitted (15-OCT-2001) Developmental Genetics, National Institute
                        TITLE      Direct Submission
                        AUTHORS    Lovell-Badge,R. and Brunelli,S.
                        REFERENCE  2  (bases 1 to 18632)
                        JOURNAL    Unpublished
                                   conserved, regulatory elements
                                   is dependent on the combined action of discrete, evolutionarily
                        TITLE      Expression of Sox3 throughout the developing central nervous system
                        AUTHORS    Brunelli,S., Bell,D., Casey,E.S., Harland,R. and Lovell-Badge,R.
                        REFERENCE  1  (bases 1 to 18632)
                        REFERENCE
                                   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        ORGANISM   Mus musculus
                        SOURCE     house mouse.
                        KEYWORDS   .
                        VERSION    AF434675.1  GI:17530192
                        ACCESSION  AF434675
                        DEFINITION Mus musculus transcription factor SOX3 (Sox3) gene, complete cds.
                        LOCUS      AF434675              18632 bp    DNA     linear   ROD 12-DEC-2001
                        AF434675
                        RESULT 11
```

```
Db    511 CGCCG 515
Qy    511 CGCCG 515
           |||  |
Db    495 CGCGG 499
Qy    495 CGCGG 499
          ||||
Db    451 GACCAAGACGCTCTCAAGAGAGGACAAGTACTCGCTGCCCGGCGCCGCGCCGCGGCCCTGGGGGGGCGCGGGGGGGGGGCGCGCAA 510
          ||| |||  |||||||| |||||||||||||||||||| |  ||| |||  || | |
Qy    435 GACCAAGACGCTGCTCAAGAGAAGGACAAGTACTCGCTGCCCGGCGTCCTGGGCGGGG 494
Db    435 GACCAAGACGCTGCTCAAGAGAAGGACAAGTACTACCCGGACCCCCCCCGCGCCCGCCAA 450
          |||||||||||||||||||||| |||||||| |  ||||  ||||||| ||||||
Qy    391 GCGACTGGCGTGCGGTGCACATGAAGGACACCCCGGAGCCGGTTCATGGTGTCCGGCGCAA 434
Db    391 GCGGCTGGCGTGCGGTGCACATGAAGGACACCCCATGAACGCCGTTCATGGTGTGGTCCCGCCGGGGGCA 390
          ||||||||||||||||||||||||||||||  |  | ||| ||||||||||| | |||   ||| ||
Qy    375 GCGGCTGCGGCGCGCGCGCGCTGCACATGAAGGAGGACACCCCGGAATTACCAAGTACCGCCCGCCGGGGGCAA 434
Db    331 GGCGCGCCACTGGAAACTGCTGACGGGGCCCGTTCATCGACGAGGCCAA 390
          |||||||||||| ||   ||||||||||||||||| ||||||| |||
Qy    315 GGGGGGCCGGAGTGGGAAGGTCATGTCCGAGGCGCCGGAGGAGGGCGCCA 374
Db    315 GGGGGGCCGGAGTGGGAAGGTCATGTCCGAGGCGCCGCGTTCATTCATCGACGAGGCCA 374
          |||||||||| ||||||||||||||||||||||||| ||  ||| |
Qy    271 GCGGGCGGCAAGGAGGATGGCCCTGGAGAACCCCAAGATGCACACAACTCCGAGATCAGCAAGCGCGGCTT 330
Db    271 GCGGGCGGCAAGGAGGATGGCCCTGGAGAACCCCAAGATGCACACAACTCGGAGATCAGCAAGCGCGGCTT 330
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Qy    255 GCGCGGCCAAGATGGCCCAAGGAGGAGAACCCCAAGAGGAGAAGGACACCCCGGAGACCCAGGAGG 314
Db    255 GCGCGGCCAAGATGGCCCAAGATGCACAACTCGGAGATCAGCAAGCGCGGCCT 314
          ||||||||||||||||||||||  |  ||   || | ||| ||| |  || |
Qy    211 CAGCGACACCAGGAGACCGCGGTCAAGGACGCCATGAACGGCGTTCATGGTGTGGTCCCGCCGGGGCCA 270
Db    211 CAGCGACACCAGGAGACCGCGGTCAAGGACGCCATGAACGGCGTTCATGGTGTGGTCCCGCCGGGGCA 270
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Qy    195 GGCCAACCAGGACGACGGGTCAAAACGGCGCCCTTCATGTGTGGTCGGCGGCCA 254
Db    195 GGCCAACCAGGACGACGGGTCAAAACGGCGCCCATGAACGGCGGGGGCGGGGGGGGGGGGCGCCA 254
          |||||||||||||||||||||| ||||||||| |  |  | ||| ||||| |||
Qy    151 CGCAGGCAACGGGGCAGCGGGGCAGCGGGGCAGCGGGGCGGGGGGGGGGGGGGCGCCCAA 210
Db    151 CGCAGGCAACGGGGCAGCGGGGCAGCGGGGCAGCGGGGCGGGGGGGGGGGGGGCGCCCAA 210
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    135 CGCCGGGGCGGCGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGCGCCGGCGCCAA 194
Db    135 CGCCGGGGCGGCGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGCGCCGGCGCCAA 194
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Matches 310; Conservative    0; Mismatches  55;    Indels     0;    Gaps     0;
Best Local Similarity    84.9%;    Pred. No. 2.1e-24;
Query Match      55.4%;      Score 277;    DB 10;    Length 1150;
ORIGIN
BASE COUNT     186 a    452 c    391 g    121 t
```

```
misc_feature              221. .457
                          /gene="sox-3"
                          /note="HMG box"

                          PPGGDAADAASPLPGGRLHGVHQHYQGAGTAVNGTVPLTHI"
                          YGQQPATAAAAAAAAAMSLGPMGSVVKSEPSSSPPAIRSHSQRACLGDLRDMISMYL
                          PGGAAAAAAAAASSPVGVGQRLDTYTHVNGWANGAYSLVQEQLGYAQPPSMSS
                          PPPPALPQMHRYDMAGLQYSPMMPPGAQSYMNAAAAAASGYGGMAPSAAAAAAA
                          LGADWKLLTDAEKRPFIDEAKRLRAVHMKEYPDYKYRPRRKTLLKKDKYSLPGGLP
                          GNGGSGGANGGGGGGGGGGSDQDRVKRPMNAFMVWSRGQRRKMALENPKMHNSEISKR
              /translation="MYSLLETELKNPVGPPTPAAGTGVPAAPGAAGKSGANPAGGANA
              /db_xref="SWISS-PROT:P53784"
              /db_xref="GI:1209432"
              /protein_id="CAA63845.1"
              /product="SOX3 protein"
              /codon_start=1
```

REFERENCE   1 (bases 1 to 855)
AUTHORS   Foster,J.W. and Graves,J.A.
TITLE   An SRY-related sequence on the marsupial X chromosome: implications
            for the evolution of the mammalian testis-determining gene
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (5), 1927-1931 (1994)
MEDLINE   94173941
REMARK   GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 145890] from the original journal article.
            This sequence comes from Fig. 2b.
FEATURES   Location/Qualifiers
source   1..855
                /organism="Sminthopsis macroura"
                /db_xref="taxon:9302"
                /chromosome="X"
gene   184..>855
                /gene="testis-determining gene"
CDS   184..>855
                /gene="testis-determining gene"
                /note="SRY homolog; This sequence comes from Fig. 2b;
                conceptual translation differs from that in published
                reference"
                /codon_start=1
                /product="testis-determining protein"
                /protein_id="AAB30154.2"
                /db_xref="GI:7579912"
                /translation="MYSLLESEMKSPVVQGPPAGGAAAPGTGKSGANAGGGGGGA
                GGGGGQDRVKRPMNAFMVWSRGQRRKMALENPKMHNSEISKRLGADWKLLTDAEKR
                PFIDEAKRLRAVHMKEYPDYKYRPRRKTKTLLKDKYSLPGNLLPPGTAAVNSPVGVG
                QRLDTYAHMNGWANGAYSLMPEQLSIYGQHPGMNGPQLQQMHRYDMAGPAVQPHDAGRP
                DLHELL"
BASE COUNT   159 a   305 c   291 g   100 t
ORIGIN

Query Match   54.6%; Score 273.2; DB 4; Length 855;
Best Local Similarity   76.5%;
Matches 335; Conservative   0; Mismatches 103; Indels   0; Gaps   0;

```
Qy      62 TGTACAGCAATGATGGAGAACCGACCTGCACTCCGCCCGCGGCGCCCCAGCGCCCCACGA 121
           | ||  ||   || | ||||||||||  ||  ||||| |||||  |||||||| || ||
Db     197 TGGAGAGTGAGATGAAAAGCCCGGTGGTCCAGGGGCCTCCGCGGCGGAGCCGGGCCCG 256
Qy     122 ACCTCTTCGGGCCCCCCCCCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCCCG 181
Db     257 CGGCCCCCCGGGGCACGGGCAAGAGCGGGCGAACCCGGCGGGGGGAGGGGGCGCCCG 316
Qy     182 GCGGGGGGGGGGGCCAAGGCCACCAAGGACCGGGGTCAAACGGCCCCATGAACGCTTTCAT 241
           |||  |||     ||||||||||||||||||| |||   ||||||||||| ||||||||
Db     317 GCGGGGGGGGGCGGCGACCAGGACGGGGTCAAGCGGCGCCCATGAACGCTTTCATTGGTGT 376
Qy     242 GGTCCCGGGGGCAGAGACGCGGGCGCCAAGGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGA 301
Db     377 GGTCCCGGGGCAGAGACGCGGAAGATGGCCCTGGAGAACCCCAAGATGCACAACTCGGAGA 436
Qy     302 TCAGCAAGCGCCTGGGCGCCGACTGGAAGCTGCTCACCGACGCCGAGAAGCGGCCGTTCA 361
Db     437 TCAGCAAGCGCCTGGGCGCCGACTGGAAGCTGCTCACCGACGCCGAGAAGCGGCCCTTCA 496
Qy     362 TCGACGAGGCCAAGCGGCTGCGCGCGGTGCACATGAAGGAGTACCCGGATTACAAGTACC 421
Db     497 TCGACGAGGCCAAGCGGCTGCGCGCGGTGCACATGAAGGAGTACCCGGACTACAAGTACC 556
Qy     422 GGCCGCGGCGCAAGACCAAGACGCTGCTCAAGAAAGACAAGTACTCGCTGCCGGGCAAGC 481
Db     557 GGCCCCGGCGCAAGACCAAGACGCTGCTCAAGGACAAGTACTCCCTGCCGGGCAACC 616
Qy     482 TCCTGGCGCGGCGGCGGCGG 499
Db     617 TGCTCCCCCGGCGGCACG 634
```

RESULT 13
LOCUS   AB011803   1824 bp   mRNA   linear   VRT 02-MAY-2001
DEFINITION Gallus gallus mRNA for SOX3, complete cds.
ACCESSION   AB011803
VERSION   AB011803.1 GI:2947026
KEYWORDS   SOX3.
SOURCE   Gallus gallus embryo lens cDNA to mRNA.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (sites)
AUTHORS   Kamachi,Y., Uchikawa,M., Collignon,J., Lovell-Badge,R. and
            Kondoh,H.
TITLE   Involvement of Sox1, 2 and 3 in the early and subsequent molecular
            events of lens induction
JOURNAL   Development 125 (13), 2521-2532 (1998)
MEDLINE   98274211
REFERENCE   2 (bases 1 to 1824)
AUTHORS   Kamachi,Y. and Kondoh,H.
TITLE   Direct Submission
JOURNAL   Submitted (03-MAR-1998) Yusuke Kamachi, Osaka University, Institute
            for Molecular and Cellular Biology; 1-3 Yamadaoka, Suita, Osaka
            565-0871, Japan (E-mail:yusuke@imcb.osaka-u.ac.jp,
            Tel:81-6-879-7964, Fax:81-6-877-1738)
FEATURES   Location/Qualifiers
source   1..1824
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /tissue_type="lens"
                /dev_stage="embryo"
gene   191..1141
                /gene="Sox3"
                /function="SRY-type HMG-domain containing transcription
                factor"
CDS   191..1141
                /gene="Sox3"
                /codon_start=1
                /product="SOX3"
                /protein_id="BAA25093.1"
                /db_xref="GI:2947027"
                /translation="MYSMLETEIKTPQPTPGSAGGNPAPGGNGKGGGGGAGSGAGSD
                QDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGADWKLLSDAEKRPFIDEAK
                RLRAVHMKEYPDYKYRPRRKTKTLLKDKYSLPGNLLAPGGNAVSSPVGVGQRIDTY
                AHMNGWTNGAYSLMDQDLGYGQHPGMNSPQLQQMHRYDMPGLQYSPMMSTAQGTYMNAA
                STYSMSPAAYGQQPSTAMSLGSMGSVVKSEPSSPPPAITSHSQRACLGDLRDMISMYL
                PPGGDADPSALQGSRLHSVHQHYQSAGTAVNGTVPLIHI"
BASE COUNT   367 a   549 c   521 g   387 t
ORIGIN

Query Match   53.5%; Score 267.6; DB 5; Length 1824;
Best Local Similarity   75.2%;
Matches 349; Conservative   0; Mismatches 109; Indels   6; Gaps   1;

```
Qy      37 CCGGTGAACCGCGACCAGCCGCGCCCCGAGTGTACAGACATGATGAGTGGAGACCGACCTGCACTCG 96
Db     164 CTGGCGCATCTCTTACCGGCGCCCCGACCATGTAACGCATGTGGAGACCGAGATCAAGACG 223
Qy      97 CCCGGCGCGGCGCCCCAGGCGCCCCACGAAACCTCTCGGGCGCCCGGGGGGCGGGGGGGGG 156
Db     224 CCCCCAGCGCCACGCCGGGC------AGCGCCGGCGGGGAACCCGGCGCCCGGGGGCAACGGC 277
Qy     157 GGCCGGAGGGCGGGGGCGGGGGGGCCAAGGCCACCAAGGACCGGGGT CAAA 216
Db     278 AAAGGCGGGCGGCGGCGGCCTTCAGTGGTCCGCGGGCGGCGTCCGGGCGGCGTGAAG 337
Qy     217 CGGCCCCATGAAGCGCCTTCATGGTGTTGGTCCCGCGGGCACCGGCCGCCAAGGATGGCCCCCAGGAG 276
Db     338 CGCCCCATGAACGCGCTTTCATGGTGTTGGTCCCGCGGGCGGCGGAAGATGGCCCCAGGAG 397
Qy     277 AACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGCGCCGAGTGGAAGGTCATG 336
Db                                                                     
```

RESULT 14

```
LOCUS       HSA327010
LOCUS       HSA327010        832 bp    DNA    linear   PRI 01-OCT-2001
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NL1-2D4R.
ACCESSION   AJ327010
VERSION     AJ327010.1    GI:15871428
KEYWORDS    
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
            Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
            Muravenko,O.V., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L.,
            Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     Analysis of NotI flanking sequences: a new tool for gene discovery
            and verification of the human genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 832)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NL1-2D4R"
BASE COUNT      175 a    273 c    269 g    114 t    1 others
ORIGIN

Query Match             52.9%;   Score 264.4;   DB 9;   Length 832;
Best Local Similarity   84.2%;   Pred. No. 6.9e-23;
Matches  298;  Conservative  0;  Mismatches  56;  Indels  0;  Gaps  0;
```

```
Qy  143 CGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGCCAAGGCCAACC 202
        ||||| |||||||||||||  ||||||||||| |||    ||||
Db   91 CGGGGGGGGGGGGGCCAACTCCACCGCGCGGGGGGGCCCAACC 150

Qy  203 AGGACCGGGTCAAACGGCCCATGAAACCGCTTCATCG...CAGCCA 262
        |||| |||||||| |||| ||||||||  |||||||| |||
Db  151 CGGACCGCGTCAAGCGCGCCCATGAAATGCCTTCATGTGTTCCC 210

Db  211 AGATGCCCCAAGAGAACCCCAAGATGCACAACTCGGAGATCAGCAA 270
Qy  263 AGGATGGCCCCAGGAGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGGCCG 322
Db  271 AGTGGGAAACTTTGTCGGAGAGAGGAGAAGCGGCCGTTCATCGACGAGCGCTAAGCGGCTGC 330
Qy  323 AGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTCATCGACGAGCGGCCAAGCGGCTGC 382
Db  331 GAGCGCGTGCACATGAAGGAGCACCCGGGATTATAAATACCGGGCCCCCGGCGGGAAAACCAAGA 390
Qy  383 GCGCGCTGCACATGAAGTGAAGGAGGACGCCCCGCGCCCCGGCGGGAAGACCAAGA 442
```

```
Db  398 AATCCCAAAGAGATGACAACTCGGAGATCAGCAAGCGCCTCGGGGCCGATTGGAAGCTGCTG 457
Qy  337 TCCGAGGCCGAGAAGCGGCCGTTCATCGAGCAGAGG...CTGCTGCCGCTGCACATG 396
Db  458 AGCGATGCCGAGAAGGGGCCCTTCATCGACGAGCAGATG...GCTGCACATG 517
Qy  397 AAGGAGGCGACCCGGATTACAAGTACCGCGGCCCCCGA...CTGCTCAAGAG 456
Db  518 AAGGAGTACCGGGATTACAAATACCGGGCCCCCGAGGAAGACCAAGACCTTGCTGAAGAAG 577
Qy  457 GACAAGTACTCGCTGCCCGGCGGGGGGGGGGGGGGCGG 500
Db  578 GACAAATACTCCCTGCCCGGGGGCCAACCTGCTGGGCCCCCCGAGGGGG 621
```

RESULT 15

```
LOCUS       AX201582
LOCUS       AX201582        1085 bp    DNA    linear   PAT 30-AUG-2001
DEFINITION  Sequence 3 from Patent WO0153349.
ACCESSION   AX201582
VERSION     AX201582.1    GI:15391430
KEYWORDS    
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1085)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE     Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 3 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US); MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US); CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..1085
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      253 a    348 c    348 g    136 t
ORIGIN

Query Match             52.9%;   Score 264.4;   DB 6;   Length 1085;
Best Local Similarity   84.2%;   Pred. No. 6.5e-23;
Matches  298;  Conservative  0;  Mismatches  56;  Indels  0;  Gaps  0;
```

```
Qy  143 CGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCCAAGGCCAACC 202
        ||||| ||||||||||||||||||||||||||||| |||   ||||
Db   65 CGGGGGGGGGGGGGGCCAACTCCACCGCGCGGGGGGGCAACCAGAAAAACAGCC 124

Qy  203 AGGACCGGGTCAAACGGCCCATGAAACCGCCTTCATGGTGTGGTCCCGGGCAGCGGGGCCA 262
        |||| |||||||| |||| ||||||||  ||||| |||||||||
Db  125 CGGACCGCGTCAAGCGCGCCCATGAAATGCCTTCATGTGTTGGTCCCGGGCACGGCGCCA 184

Db  185 AGATGGCCCCAAGACGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCCTGGGCGCCG 244
Qy  263 AGGATGGCCCCAGGAGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGGCCG 322
Db  245 AGTGGGAAACTTTGTCGGAGAGAGGAGAAGCGGCCGTTCATCGACGAGCGCTAAGCGGCTGC 304
Qy  323 AGTGGAAGGTCATGTCCGAGGCCGAGAAGCGGCCGTTCATCGACGAGCGGCCAAGCGGCTGC 382
Db  305 GAGCGCGTGCACATGAAGGAGCACCCGGGATTATAAATACCGGGCCCCCGGCGGGAAAACCAAGA 364
Qy  443 CGCTGCTCAAGAGAGGACGAAGAGTACTCCGCTGGGCCGCCGGGCGGGCCGG 496
Db  365 CGCTCATGAAGAAGGATAAGTACTACAGGCTGCCCGGGCTGCCTGGGGCCCCGGGCCG 418
```

```
Qy  443 CGGTGCTCAAGAGAGGACGAAGAGTACTCCGCTGGGCCGCCGGGCGGGCCGG 496
        |||| |    |||| ||||   ||||  ||     ||||||||||||  ||||
Db  391 CGCTCATGAAGAAGGATAAGTACTACACGGTTGCCGTGCCTGCTTGGGGCGG 444
```

```
Search completed: October 10, 2002, 17:30:13
Job time : 799.867 secs
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 seconds
(without alignments)
14616.225 Million cell updates/sec

Title: US-09-489-101A-5_COPY_1_500
Perfect score: 500
Sequence: 1 atgctcctggacgcgggtcc..........ccgggcctgccaagacagca 500

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326829 residues
Total number of hits satisfying chosen parameters: 3595312

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database : GenEmbl:*
   1: gb_ba:*
   2: gb_htg:*
   3: gb_in:*
   4: gb_om:*
   5: gb_ov:*
   6: gb_pat:*
   7: gb_ph:*
   8: gb_pl:*
   9: gb_pr:*
  10: gb_ro:*
  11: gb_sts:*
  12: gb_sy:*
  13: gb_un:*
  14: gb_vi:*
  15: gb_ba:*
  16: gb_fun:*
  17: em_hum:*
  18: em_in:*
  19: em_mu:*
  20: em_om:*
  21: em_or:*
  22: em_ov:*
  23: em_pat:*
  24: em_ph:*
  25: em_pl:*
  26: em_ro:*
  27: em_sts:*
  28: em_un:*
  29: em_vi:*
  30: em_htg_hum:*
  31: em_htg_inv:*
  32: em_htg_other:*
  33: em_htg_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	1602	6	AX201584	AX201584 Sequence
2	500	100.0	1602	6	AX321400	AX321400 Sequence
3	500	100.0	851	9	HSA322705	AJ322705 Homo sapi
4c	467.6	93.5	2680	9	AF104902	AF104902 Homo sapi
5c	467.6	93.5	153762	9	AL355338	AL355338 Human DNA
6	376.8	75.4	2444	10	MUSZIC2P1	D70848 Mouse mRNA
7c	169.8	34.0	851	9	HSA322705	AJ322705 Homo sapi
8c	130.2	26.0	42999	9	AC104208	AC104208 Mus muscu
9c	114.8	23.0	1255	9	AF104902	AF104902 Homo sapi
10c	114.2	23.0	174559	9	AC027060	AC027060 Homo sapi
11	113.2	22.6	3138	9	AR103244	AR103244 Sequence
12	113.2	22.6	3138	8	BD000107	BD000107 Different
13	113.2	22.6	3138	9	HUM2ICP	D76435 Homo sapien
14	106.8	21.4	59150	2	AC095557	AC095557 Rattus no
15	105.2	21.0	1344	6	AF221839	AF221839 Rattus no
16	105.2	21.0	2947	9	AX305309	AX305309 Sequence
17	105.2	21.0	2947	10	D32167	D32167 Mouse zic m
18	81	16.2	3132	5	MUSZIC	MUSZIC Mouse zic
19c	73.2	14.6	176370	2	AF151535	AF151535 Danio rer
20	72.6	14.5	2239	5	AC105829	AC105829 Rattus no
21	70.2	14.0	1825	5	AF207751	AF207751 Danio rer
22	70.2	14.0	1838	5	AB009564	AB009564 Xenopus l
23	70.2	14.0	2435	5	AF022927	AF022927 Xenopus l
24	69.2	13.8	63082	2	AF028805	AF028805 Xenopus l
25	68.4	13.7	159716	2	AC022663	AC022663 Homo sapi
26	67.8	13.6	52884	2	AC103234	AC103234 Rattus no
27	67.2	13.4	161307	2	AC068018	AC068018 Homo sapi
28	66.4	13.3	3444	9	AC084058	AC084058 Homo sapi
29	66.4	13.3	103978	9	AF028706	AF028706 Homo sapi
30c	66.4	13.3	109339	9	AC106628	AC106628 Rattus no
31	66.4	13.3	176370	2	AL035443	AL035443 Human DNA
32	66.2	13.3	201430	5	AC105829	AC105829 Rattus no
33	65.2	13.0	2901	5	HS137H15	Z96804 Homo sapien
34	64.2	12.8	3509	10	AB009565	AB009565 Xenopus l
35	64.2	12.8	208531	2	MUSZIC3P2	D70849 Mouse mRNA
36	63.6	12.7	153170	2	AC087560	AC087560 Mus muscu
37	63.6	12.7	113431	2	AC103491	AC103491 Rattus no
38	63	12.6	208936	2	AC094306	AC094306 Rattus no
39	62.8	12.6	38426	1	AC010821	AC010821 Homo sapi
40	62.8	12.6	16274	2	SCF55	AL132991 Streptomy
41c	62.6	12.5	99894	2	AC110453	AC110453 Rattus no
42	62.2	12.4	82985	2	AC096067	AC096067 Rattus no
43	62.2	12.4	98421	2	AC105572	AC105572 Rattus no
44	62	12.4	131728	2	AC105896	AC105896 Rattus no
45	61.8	12.4	125020	9	AC105719	AC105719 Rattus no
					AF429315	AF429315 Homo sapi

ALIGNMENTS

RESULT 1
AX201584

LOCUS   AX201584   1602 bp   DNA   linear   PAT 30-AUG-2001
DEFINITION Sequence 5 from Patent WO0153349.
ACCESSION   AX201584
VERSION   AX201584.1 GI:15391433
KEYWORDS
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1602)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE   Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 5 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES   Location/Qualifiers
     source   1..1602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

```
BASE COUNT     275 a      578 c      536 g      213 t
ORIGIN

Query Match
Best Local Similarity  100.0%;
Matches 500;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 ATGCTCCTGGACGCGGGGTCCGCAGTTCCCGGCGCCATCGGGGTGGGGCAGCTTCGCGGCCAC   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCTCCTGGACGCGGGGTCCGCAGTTCCCGGCGCCATCGGGGTGGGGCAGCTTCGCGGCCAC   60
```

RESULT 2

```
LOCUS       AX321400               1602 bp    DNA     linear   PAT 15-DEC-2001
DEFINITION  Sequence 417 from Patent WO0177168.
ACCESSION   AX321400
VERSION     AX321400.1  GI:17905325
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Lodes,M.J., Wang,T., Mohamath,R. and Indrias,C.Y.
TITLE       Compositions and methods for the therapy and diagnosis of lung
            cancer
JOURNAL     Patent: WO 0177168-A 417 18-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..1602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
BASE COUNT     275 a      578 c      536 g      213 t
ORIGIN

Query Match
Best Local Similarity  100.0%;
Matches 500;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 ATGCTCCTGGACGCGGGGTCCGCAGTTCCCGGCGCCATCGGGGTGGGGCAGCTTCGCGGCCAC   60
```

RESULT 3

```
LOCUS       AF104902               1602 bp    mRNA    linear   PRI 17-DEC-1998
DEFINITION  Homo sapiens ZIC2 protein (ZIC2) mRNA, complete cds.
ACCESSION   AF104902
VERSION     AF104902.1  GI:4028591
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1602)
AUTHORS     Brown,S.A., Warburton,D., Brown,L.Y., Yu,C.Y., Roeder,E.R.,
            Stengel-Rutkowski,S., Hennekam,R.C. and Muenke,M.
TITLE       Holoprosencephaly due to mutations in ZIC2, a homologue of
            Drosophila odd-paired
JOURNAL     Nat. Genet. 20 (2), 180-183 (1998)
MEDLINE     98444655
REFERENCE   2 (bases 1 to 1602)
AUTHORS     Brown,S., Brown,L.Y. and Warburton,D.
TITLE       Direct Submission
JOURNAL     Submitted (06-NOV-1998) Obstetrics and Gynecology, Columbia
            University, 630 W. 168th St., New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1..1602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q32"
     gene            1..1602
                     /gene="ZIC2"
     CDS             1..1602
                     /gene="ZIC2"
                     /codon_start=1
                     /product="ZIC2 protein"
                     /protein_id="AAC96325.1"
                     /db_xref="GI:4028592"
```

Left record

REFERENCE   2   (bases 1 to 2680)
  PUBMED   1098499
  JOURNAL   J. Biol. Chem. 275 (49), 38863-38869 (2000)
  TITLE   ZIC2 and Sp3 Repress Sp1-induced Activation of the Human D1A
            Dopamine Receptor Gene
  AUTHORS   Yang,Y.; Hwang,C.K.; Junn,E.; Lee,G. and Mouradian,M.M.
REFERENCE   1   (bases 1 to 2680)
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  ORGANISM Homo sapiens
  SOURCE   human.
  KEYWORDS .
  VERSION   AF193855.1 GI:11065969
  ACCESSION AF193855
  DEFINITION Homo sapiens zinc finger protein of cerebellum ZIC2 (ZIC2) mRNA,
            complete cds.
  LOCUS   AF193855   2680 bp   mRNA   linear   PRI 05-DEC-2000
RESULT 4

                     /translation="MLLDAGPQFPAIGVGSFARHHHHSAAAAAAAAEMQDRELSLAA
                     AQNGFVDSAAAHMGAFKLNPGAHELSPGQSSAAFTSQGPGAYPGSAAAAAAALGPHA
                     AHVGSYSGPPFNSTRDFLPRSARLPGTSAPGGQHGLFGPGAGGLHAHSDAQGHLLF
                     PGLPEQHGPHGSQNVLNGQMRLGLPGEVFGRGSEQYRQVASPRTDPYSAAQLHNQYGPM
                     NMNMGMNMAAAAHHHHHHHPGAFFRYMRQQCIKQELICKWIDPEQLSNPKKSCNK
                     TFSTMHELYTHVSVEHVGGPEQSNHVCFWEECPREGKPFKAXYKLVNHIRVHTGEKPF
                     PCPFPGCGKVFARSENLKIHKRTHTGEKPFQCEFEGCDRRFANSSDRKKHMHVHTSDK
                     PYLCKMCDKSYTHPSSLRKIHMKVHHSSPQGSESSPAASSGYESSSTPPGLVSPSAEPQS
                     SSNLSPAAAAAAAAAAAAAVSAVHRGGSGSGSGSGSGSGSGSGSGSGGSGGGGGG
                     GSSGGGSGTAGGHSGLSSNFENWYV"

BASE COUNT   275 a   578 c   536 g   213 t
ORIGIN

Query Match   100.0%; Score 500; DB 9; Length 1602;
Best Local Similarity   100.0%; Pred. No. 7.2e-51; Mismatches 0; Indels 0; Gaps 0;
Matches 500; Conservative 0;

Qy   1 ATGCTCCTGGACGCGGGTCCCGCAGTTCCCGCAGTTCGGGGTGGGTGGGCAGCTTCGCGCCGCCAC   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCTCCTGGACGCGGGGTCCCGCAGTTCCCGGCAGTTCCGCGCCGCCAC   60

Qy   61 CATCACCACTCCGCGCCGCCCGGGCGGTGCCCGCGAGATGCAGGACCGTGAACTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATCACCACTCCGCGCCGCCCGGGCGGTGCCCGCGAGATGCAGGACCGTGAACTG 120

Qy 121 AGCCTGGCGGCGCGCAGAACGGCGCTTCGTTGTTGATTCCGCGCCACATGGGAGACCTTC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGCCTGGCGGCGCGCAGAACGGCTTCGTTGATTCCGCGCCACATGGGAGACCTTC 180

Qy 181 AAGCTCAACCCGGGCGCCCACGAGCTGTGCCCGGGCCCAGAGCTGTTCACGTCGCAG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AAGCTCAACCCGGGCGCCCACGAGCTGTGCCCGGGCCCAGAGCTGTTCACGTCGCAG 240

Qy 241 GGCCCCGGGCGGCCTACCCCGGGCTCCGCGCCGCCGCCGCCCCGGGGGACTTCCTG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GGCCCCGGGCGGCCTACCCCGGGCTCCGCGCCGCCGCCGCCCCGGGGGACTTCCTG 300

Qy 301 CACGGCCGCGCACGTTGGCGTCCTACTCTCGGGGACTTCAAACTCCACCCGGGGACTTCCTG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 CACGGCCGCGCACGTTGGCGTCCTACTCTCGGGGACTTCAAACTCCACCCGGGGACTTCCTG 360

Qy 361 TTCCCGCACCGCGCGCGGGCGTTCGGGCGCGCAGCAGCGGCGCACGGGGTCGTGTTC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 TTCCCGCACCGCGCGCGGGCGTTCGGGCGCGCAGCAGCGGCGCACGGGGTCGTGTTC 420

Qy 421 GGGCCGCGGGCGGGCGGCCGGCCGCTGCCACCACGCGCCACCGGGGCTGTTC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GGGCCGCGGGCGGGCGGCCGGCCGCTGCCACCACGCGCCACCGGGGCTGTTC 480

Qy 481 CCGGGGCCTGCCAGAGCAGCA 500
        |||||||||||||||||||||
Db 481 CCGGGGCCTGCCAGAGCAGCA 500

Right record

  AUTHORS   Yang,Y. and Mouradian,M.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-OCT-1999) NINDS, NIH, 10 Center Drive, Bethesda, MD
            20892, USA
FEATURES   Location/Qualifiers
     source   1..2680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene   1..2680
                     /gene="ZIC2"
     CDS   1..1599
                     /gene="ZIC2"
                     /note="transcription factor; involved in cerebellar
                     development"
                     /codon_start=1
                     /product="zinc finger protein of cerebellum ZIC2"
                     /protein_id="AAG28409.1"
                     /db_xref="GI:11065970"
                     /translation="MLLDAGPQFPAIGVGSFARHHHHSAAAAAAAAEMQDRELSLAA
                     AQNGFVDSAAAHMGAFKLNPGAHELSPGQSSAAFTSQGPGAYPGSAAAAAAALGPHA
                     AHVGSYSGPPFNSTRDFLPRSRGFGDSAPGGQHGLFGPGAGGLHAHSDAQGHLLFP
                     GLPEQHGPHGSQNVLNGQMRLGLPGEVFGRGSEQYRQVASPRTDPYSAAQLHNQYGPM
                     MNMGMNMAAAAHHHHHHHPGAFFRYMRQQCIKQELICKWIDPEQLSNPKKSCNKT
                     FSTMHELYTHVSVEHVGPEQSNHVCFWEECPREGKPFKAYKYLVNHIRVHTGEKPF
                     CPFPGCGKVFARSENLKIHKRTHTGEKPFQCEFEGCDRRFANSSDRKKHMHVHTSDKP
                     YLCKMCDKSYTHPSSLRKHMKVHHSSPQGSESSPAASSYESSSTPPGLVSPSAEPQSS
                     SNLSPAAAAAAAAAAAVSAVHRGGSGSGSGSGSGSGSGSGSGSGGSGGGGGG
                     SSGGGSGTAGGHSGLSSNFNEWYV"

BASE COUNT   572 a   798 c   723 g   586 t   1 others
ORIGIN

Query Match   93.5%; Score 467.6; DB 9; Length 2680;
Best Local Similarity   98.6%; Pred. No. 4.2e-47; Indels 3; Gaps 2;
Matches 493; Conservative 0; Mismatches 4;

Qy   1 ATGCTCCTGGACGCGGGTCCCGCAGTTCCCGCAGTTCCCGCCCATGGGGTGGGTGGGCAGC   60
        ||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   1 ATGCTCCTGGACGCGGGGTCCCGCAGTTCCCGGCAGTTCCCGCGCCAC   60

Qy   61 CATCACCACTCCGCGCCGCCCGGGCGGTGCCCGCGAGATGCAGGACCGTGAACTG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATCACCACTCCGCGCCGCCCGGGCGGTGCCCGCGAGATGCAGGACCGTGAACTG 120

Qy 121 AGCCTGGCGGCGCGCAGAACGGCGCTTCGTTGATTCGTTGATTCCGCGCCACATGGGAGACCTTC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGCCTGGCGGCGCGCAGAACGGCTTCGTTGACTCCGCGCCACATGGGAGACCTTC 180

Qy 181 AAGCTCAACCCGGGCGCGCACGAGCGTGTCCCGGGCCCAGAGCTGTTCACGTCGCAG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AAGCTCAACCCGGGCGCGCACGAGCGTGTCCCGGGCCCAGAGCTGTTCACGTCGCAG 240

Qy 241 GGCCCCGGGCGGCCTACCCCGGGCTCCGCGCCGCCGCCGCCCCGGGGGACTTCCTG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GGCCCCGGGCGGCCTACCCCGGGCTCCGCGCCGCCGCCGCCCCGGGGGACTTCCTG 300

Qy 301 CACGGCCGCGCACGTTGGCTCCTACTCTCGGGGACTTCAAACTCCACCCGGGGACTTCCTG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 CACGCCCGCGCACGTTGGCTCCTACTCTCGGGGACTTCAAACTCCACCCGGGGACTTCCTG 360

Qy 361 TTCCCGCAGCGCGCGCGGGCGCTTCCGGGGGACTTCCGGGGGCGGCAGCGGCGCACGGGGGTCGTGTTC 420
        ||||||||||||| |||| |||
Db 361 TTCCCGCAGC--CCCGGCTTC--GGGGACTCGCGCCGGGCGGCGAGCAGCGACGGCGGTCGTGTTC 417

Qy 421 GGGCCGCGGGCGGGCGGCCGCTGCACCACGCGCCACCTCTCTCTTC 480
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 418 GGGCCGCGGGCGGGCGGCCTGCACCACGCGCACTCGGACACCTCTCTCTTC 477

Qy 481 CCGGGGCCTGCCAGAGCAGCA 500
        |||||||||||||||||||||
Db 478 CCGGGGCCTGCCAGAGCAGCA 497

RESULT 5

```
LOCUS       AL355338              153762 bp    DNA     linear   PRI 25-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-12G12 on chromosome 13, complete
            sequence.
ACCESSION   AL355338 AC011573
VERSION     AL355338.33  GI:15787728
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153762)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 26, 2001 this sequence version replaced gi:15590915.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-12G12 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-12G12 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-340C20 is at 151763 in this
            sequence. The true right end of clone RP11-134O15 is at 2000 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1..153762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-12G12"
                     /clone_lib="RPCI-11.1"
     misc_feature    16173..17736
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    1913..20328
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    2829..32398
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    3148..31554
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    39803..45993
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
                     /note="CpG island"
     misc_feature    41928..42105
                     /evidence=not_experimental
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    49260..49278
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly. Sequence
                     from uni-directional primer reads and dGTP big dye
                     terminator reads only."
     misc_feature    49385..50238
                     /note="CpG island"
     misc_feature    57373..58438
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    66327..66439
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    110145..110430
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    11420..114301
                     /note="match: STS: Em:G17681"
     misc_feature    149007..149855
                     /note="match: STS: Em:G20811"
                     /note="CpG island"
     misc_feature
                     /evidence=not_experimental

BASE COUNT    41323 a  35024 c  35935 g  41480 t
ORIGIN
```

Query Match   93.5%;   Score 467.6;   DB 9;   Length 153762;
Best Local Similarity   98.6%;   Pred. No. 1.2e-47;
Matches   493;   Conservative   0;   Mismatches   4;   Indels   3;   Gaps   2;

```
Qy          1 ATGCTCCTGGACGGCGGGGTCCCGGCCAGTTCCCGGCCCATCGGGGTGGGCAGCTTCGCGCCAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42369 ATGCTCCTGGACGGCGGGGTCCGCAGTTCCCGGCCCATCGGGGTGGGCAGCTTCGCGCGCCAC 42428

Qy         61 CATCACCAGTTCCGCCGCGCGGGGCGGGGCGGCGGCTGCCGCCCCAGAATGCAGGACGTGAACTG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42429 CATCACCAGTTCCGCCGCGCGGGGCGGGGCGGCGGCTGCCGCCCCAGAATGCAGGACGTGAACTG 42488

Qy        121 AGCCTGGCGCGGCGGCCAGAACGGCGCTTCGTTGATTCCGCCGCCGCGCCACCATGGGAGGCCTTC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42489 AGCCTGGCGCGGCGGCCAGAACGGCTTCGTTGACTCCGCCGCCGCGCCACATGGGAGGCCTTC 42548

Qy        181 AAGCTCAACCCGGCGCCACGGAGCTGTGCCCCGGGCCAGAGCTCGGCGTTCACGTCGCAG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42549 AAGCTCAACCCGGCGCCACGGACCTGTCCCCGGGCCAGAGCTCGGCGTTCACGTCGCAG 42608

Qy        241 GGCCCGGCGCCGCCTACCCGGCTCCGCTGCGGCGTCGGGCGCGGCGCCTCGGGCGCCC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42609 GGCCCGGCGCCGCCTACCCGGGCTCCGCTGCGGCGTCGGGCGCGGCGCCTCGGGCGCCC 42668

Qy        301 CACGCCCGCGCACGTTGGCTCCTACTCTGGGCGCCCTTCAACTCCACCCGGGACTTCCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42669 CACGCCCGCGCACGTTGGCTCCTACTCTGGGCGCCCTTCAACTCCACCCGGGACTTCCTG 42728

Qy        361 TTCCCGCAGCGCGCGCGGGCTTCGGCGCGCGGGCGCGGGGGCGCACGCACGGGCTGTGTTC 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      42729 TTCCGCAC-CGCGGCTTC--GGGGACTCGGCGGCGCGGGCGCGGGGGCGCAGCAGGGGCTGTGTTC 42785

Qy        421 GGGCCGGCGGGGCGGGGCGCTGACACCGGCGACCGGCGCAGGGCGCAGCAGCACGGGCTGTTC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42786 GGGCCGGCGGGGCGGGGCGGGGCTGCGACCACCACGGCGCACTCGGACGCGCAGGGCCACCTCTCTTC 42845

Qy        481 CCGGGGCCTGCCAGAGCACA 500
              ||||||||||||||||||||
Db      42846 CCGGGGCCTGCCAGAGCACA 42865
```

RESULT 6

```
LOCUS       MUS2IC2P1               2444 bp    mRNA    linear   ROD 03-FEB-1999
DEFINITION  Mouse mRNA for zic2 protein, complete cds.
ACCESSION   D70848
VERSION     D70848.1  GI:1345412
KEYWORDS    zic2 protein.
SOURCE      Mus musculus cerebellum cDNA to mRNA.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2444)
AUTHORS     Aruga,J.
TITLE       Direct Submission
JOURNAL     Submitted (04-OCT-1995) Jun Aruga, Institute of Physical and
            Chemical Research (RIKEN), Molecular Neurobiology Laboratory; 3-1-1
            Koyadai, Tsukuba, Ibaraki 305, Japan (Tel:0298-36-9170,
            Fax:0298-36-9040)
REFERENCE   2  (bases 1 to 2444)
AUTHORS     Aruga,J.
JOURNAL     Unpublished (1996)
REFERENCE   3  (sites)
AUTHORS     Aruga,J., Nagai,T., Tokuyama,T., Hayashizaki,Y., Okazaki,Y.,
            Chapman,V.M. and Mikoshiba,K.
TITLE       The mouse zic gene family. Homologues of the Drosophila pair-rule
            gene odd-paired
JOURNAL     J. Biol. Chem. 271 (2), 1043-1047 (1996)
MEDLINE     96132843
FEATURES             Location/Qualifiers
     source          1..2444
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="cerebellum"
                     /chromosome="14"
     gene            243..1835
                     /gene="zic2"
     CDS             243..1835
                     /gene="zic2"
                     /codon_start=1
                     /product="zic2 protein"
                     /protein_id="BAA11115.1"
                     /db_xref="GI:1345413"
                     /translation="MLLDAQPQFPAIGVGSFARHHHHSAAAAAAAAEMQDRELSLAA
                     AQNGFVDSAAAHMGAFKLESPGQSSAFTSQGPGAYIPGSAAAAAAAAALGPHA
                     AHVGSYSGGPPFNSTRDFLFRSRGFGDSAPGGGQHGLPGPGAGGLHHAHSDAQGHLLPH
                     GLPEQEOHGPHASQNVLNGQMRLGLPGEVFGVRSEQYRQVASPRTDYSAAQLHNQYGPM
                     NMNQMNMAAAAHHHHHHHPGAFPYMYRQQCIKQELICKWIDPEQLSNPKKSCNKT
                     FSTMHELVTHVSVEHVGGPEQSNHVCFWEECPREGKPFKAKYKLVNHIRVHTGEKPFP
                     CPFPGCGKVFARSENLKIHKRTHTEKPFQCEFEGCDRRFANSSDRKKHMHVHTSDKP
                     YLCKMCDKSYTHPSSLRKHMKVHESSPQGSESSPAASSGYESSTPPGLVSPSAEPQSS
                     SNLSPAAAAAAAAAARAVSAVHRGAGSGSSGGARRPAAVGAAAGRAAGAAAAL
                     AGAAGTTGHSGLSSNFNEWYV"
BASE COUNT      484 a    779 c    688 g    493 t
ORIGIN
```

Query Match   75.4%; Score 376.8; DB 10; Length 2444;
Best Local Similarity   86.08%; Pred. No. 2.4e-36;
Matches 430; Conservative 0; Mismatches 67; Indels 3; Gaps 1;

```
Qy      1 ATGGCTCCTGGACGCGCGGGTCCCGAGTTCCCGGCGCGCTCCCGGCGCGCTCCCGGGGCCC  60
          |||| ||| |||||||||||||||| ||||||||| ||||||| |||||||||||| | |||
Db    243 ATGCTTCTTGGACGCGCGGGTCCCGCGTTCCCGGCCCATCGGGTGGGGTGGGGCAGCTTCC  302

Qy     61 CATCCACCACTCCGCCCCGGCCCCCGGCCCCCGAGATGCAGGACCGTGAACTG         120
          |||||||||||| |||||| || | |||| |||| ||| |||| ||  |  |||
Db    303 CACCACCACTCGGCCCCGGCGTGCGGGCTGCGGGGCAGGAGGACCGCGAGCTG         362

Qy    121 AGCCTGGCGGCGGCGCAGAACGGCTTCGTTGATTCCGCGGCGCCACACATGGGAGCCTTC  180
          |||||| || | ||||||||| | |||||||| ||||||| ||  | |||  | || |||
Db    363 AGCCTGCGCGGCGCTCAGAACCGGCTTCGTGGACTCGGCGGCGCACATGGGCGCCTTC   422

Qy    181 AAGCTCAACCCCGGCGCGGCGCCAGGAGCGTGTCCCCGGCGTTCACGTCCCAG         240
          |||||||||  ||||| ||||||| | || |||| || |  |||||||| ||
Db    423 AAGCTCAACACCCGGCGGCGCCACGACGAACGTGTCTCCTGGTCAGAGTCTCGGCGTTCACGTCCCAA  482
```

```
RESULT 7
HSA322705/c
LOCUS       HSA322705                851 bp    DNA     linear   PRI 01-OCT-2001
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NR1-OD1C.
ACCESSION   AJ322705
VERSION     AJ322705.1  GI:15867084
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 851)
AUTHORS     Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
            Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
            Miravenko,O.V., Protopopov,A.I., Kashuba,V.I. Kisselev,L.L.,
            Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
TITLE       Analysis of NotI flanking sequences: a new tool for gene discovery
            and verification of the human genome
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 851)
AUTHORS     Zabarovsky,E.R.
TITLE       Direct Submission
JOURNAL     Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..851
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NR1-OD1C"
BASE COUNT      175 a    222 c    223 g    231 t
ORIGIN
```

Query Match   34.0%; Score 169.8; DB 9; Length 851;
Best Local Similarity   80.8%; Pred. No. 1e-11;
Matches 235; Conservative 0; Mismatches 52; Indels 4; Gaps 3;

```
Qy      1 ATGGCTCCTGACGCGCGGGTCG--CAGTTCCCGGCCATCGGGTGGGGTGGGGCAGCTTCGC-GCGC  57
          |||| |||| |||| |||| ||   |||||||| ||||||||| ||||||||||| |  ||  ||||
Db    291 ATGCTCCTGTACGTGGGTCCGGCCTCCCCGACAGAATCCTTGGCCATCGGGGTGGGGCATCTTCGCGAGAGC  232

Qy     58 CACCATCACCACTCCGCCCCGCCCGCCGCGCGCGGCGGCGGGAATGCAGGAACCGTGAA  117
          |||||||||||| ||||||  ||  |||||| ||||| ||| ||  |||||   | ||
Db    231 CACCATCACCACGTCGCCCCGCGGGTCGGGCGTGGGCGGGTGCCGGAGGACGCAAGGCC  172

Qy    118 CTG-AGCCTGGCGGCGGCGCCAGGAGACGGCTTCGTTCGTTGATTCCGCGGCGCACATGGGAGC  176
          ||| ||||||  | | ||||  | ||  |||||||  |||| ||||| ||   |  ||| || ||
Db    171 ATGTATCCTGTAGGTTGGGCGCAGAACGGCTTCGTAGAAATCCTTCGCCGAGAGCGCACATGGTAGC  112

Qy    177 CTTCAAGCTCAACCCCGGCGCGGCGCCACGAGCTGTCCCCGGCGTTCACGTC  236
```

```
RESULT 8

Db  111  ATTCAAGCTCAAGCCGGGCCGCCGCTACCCGCGGCTCCGGTGCCCGTGAGCTGTCCGCGGGGCCGCAGAGCTCGGAGTTAACGTT
         |||||||||| |||| |||| |||| |||| |||||||||||||||| |||||||| |||||||| |||| |||||||||| |||| ||||
Qy  237  GCAGGGGCCCGGGCCGCCGCTACCCGCGGCTCCGGTGCCGGGGCCGCAGAGCTCGGAGTTAACGTT 52
         |||||||||| |||| |||||||||||||| |||||||||||||||| |||||||| ||||||| |||||
Db   51  GCAGGTCCCGGCGGCGCCAACCCGGCTCATAAGCGGTGCCACAGCGGCCGC 1
```

```
AC104208/c

LOCUS       AC104208               42999 bp    DNA     linear   HTG 06-DEC-2001
DEFINITION  Mus musculus clone RP23-14204, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC104208
VERSION     AC104208.1  GI:17386373
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 42999)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-14204
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42999)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camaratta,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,M., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,N., Theodore,J.,
            Topham,K., Travis,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------------------------------------
                    Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------------------------------------
                    Project Information
                Center project name: L19633
                Center clone name: 142_O_4
            ---------------------------------------
            * NOTE: This record contains 55 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
```

```
*       1    680: contig of 680 bp in length

*     681    780: contig of 100 bp in length
*     781   1476: contig of 696 bp in length
*    1477   1576: gap of 100 bp
*    1577   2267: contig of 691 bp in length
*    2268   2367: gap of 100 bp
*    2368   3068: contig of 701 bp in length
*    3069   3168: gap of 100 bp
*    3169   3854: contig of 686 bp in length
*    3855   3954: gap of 100 bp
*    3955   4622: contig of 668 bp in length
*    4623   4722: gap of 100 bp
*    4723   5425: contig of 703 bp in length
*    5426   5525: gap of 100 bp
*    5526   6198: contig of 673 bp in length
*    6199   6298: gap of 100 bp
*    6299   6964: contig of 666 bp in length
*    6965   7064: gap of 100 bp
*    7065   7745: contig of 681 bp in length
*    7746   7845: gap of 100 bp
*    7846   8529: contig of 684 bp in length
*    8530   8629: gap of 100 bp
*    8630   9326: contig of 697 bp in length
*    9327   9426: gap of 100 bp
*    9427  10112: contig of 686 bp in length
*   10113  10212: gap of 100 bp
*   10213  10873: contig of 661 bp in length
*   10874  10973: gap of 100 bp
*   10974  11604: contig of 631 bp in length
*   11605  11704: gap of 100 bp
*   11705  12377: contig of 673 bp in length
*   12378  12477: gap of 100 bp
*   12478  13147: contig of 670 bp in length
*   13148  13247: gap of 100 bp
*   13248  13921: contig of 674 bp in length
*   13922  14021: gap of 100 bp
*   14022  14715: contig of 694 bp in length
*   14716  14815: gap of 100 bp
*   14816  15511: contig of 696 bp in length
*   15512  15611: gap of 100 bp
*   15612  16295: contig of 684 bp in length
*   16296  16395: gap of 100 bp
*   16396  17070: contig of 675 bp in length
*   17071  17170: gap of 100 bp
*   17171  17850: contig of 680 bp in length
*   17851  17950: gap of 100 bp
*   17951  18610: contig of 660 bp in length
*   18611  18710: gap of 100 bp
*   18711  19375: contig of 665 bp in length
*   19376  19475: gap of 100 bp
*   19476  20139: contig of 664 bp in length
*   20140  20239: gap of 100 bp
*   20240  20928: contig of 689 bp in length
*   20929  21028: gap of 100 bp
*   21029  21709: contig of 681 bp in length
*   21710  21809: gap of 100 bp
*   21810  22479: contig of 670 bp in length
*   22480  22579: gap of 100 bp
*   22580  23248: contig of 669 bp in length
*   23249  23348: gap of 100 bp
*   23349  24054: contig of 705 bp in length
*   24054  24153: gap of 100 bp
*   24154  24843: contig of 690 bp in length
*   24844  24943: gap of 100 bp
*   24944  25637: contig of 694 bp in length
*   25638  25737: gap of 100 bp
*   25738  26436: contig of 699 bp in length
*   26437  26536: gap of 100 bp
*   26537  27241: contig of 705 bp in length
*   27242  27341: gap of 100 bp
*   27342  28005: contig of 664 bp in length
*   28006  28105: gap of 100 bp
*   28106  28779: contig of 674 bp in length
*   28780  28879: gap of 100 bp
```

```
FEATURES             Location/Qualifiers
     source          1. .42999
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-14204"
                     /clone_lib="RPCI-23 Female Mouse BAC"

     *    28880  29548: contig of 669 bp in length
     *    29549  29648: gap of 100 bp
     *           30335: contig of 687 bp in length
     *    30336  30435: gap of 100 bp
     *    30436  31117: contig of 682 bp in length
     *    31118  31217: gap of 100 bp
     *    31218  31912: contig of 695 bp in length
     *    31913  32012: gap of 100 bp
     *    32013  32662: contig of 650 bp in length
     *    32663  32762: gap of 100 bp
     *    32763  33448: contig of 686 bp in length
     *    33449  33548: gap of 100 bp
     *    33549  34247: contig of 699 bp in length
     *    34248  34347: gap of 100 bp
     *    34348  35038: contig of 691 bp in length
     *    35039  35138: gap of 100 bp
     *    35139  35829: contig of 691 bp in length
     *    35830  35929: gap of 100 bp
     *    35930  36612: contig of 683 bp in length
     *    36613  36712: gap of 100 bp
     *    36713  37402: contig of 690 bp in length
     *    37403  37502: gap of 100 bp
     *    37503  38193: contig of 691 bp in length
     *    38194  38293: gap of 100 bp
     *    38294  39052: contig of 759 bp in length
     *    39053  39152: gap of 100 bp
     *    39153  39851: contig of 699 bp in length
     *    39852  39951: gap of 100 bp
     *    39952  40644: contig of 693 bp in length
     *    40645  40744: gap of 100 bp
     *    40745  41433: contig of 689 bp in length
     *    41434  41533: gap of 100 bp
     *    41534  42211: contig of 678 bp in length
     *    42212  42311: gap of 100 bp
     *    42312  42999: contig of 688 bp in length.

BASE COUNT    10052 a    8136 c    8694 g    10628 t    5489 others
ORIGIN

Query Match 26.0%;  Score 130.2;  DB 2;  Length 42999;
Best Local Similarity 55.0%;  Pred. No. 1.4e-07;
Matches 160;  Conservative 0;  Mismatches 128;  Indels 3;  Gaps 1;

Qy     210 CCCGGGGCCAGAGCTCGGCGTTCACGTNCGCAGGGCCCCGGGCCTACCCGGCGTCCGCTGC 269
           ||
Db   22584 CCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 22525

Qy     270 GGGCTGCGCGTCGGCGCCGCAGCCGCTCGGGCCCCCACGCGTTGGCGTCTACTCTCTGG 329
Db   22524 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGGCTCC 22465

Qy     330 GCCGCGCCTTCAACTCCAACCCGGGGACTTCCTGTTCCGCAGCGCGGCGGCGGCGGACTTC 389
Db   22464 GCCTCCCTTTAATTCACCCGGGGGACTTCCGTGTTCCCGCAGC---CGGGGCGTTCGGGGA 22408

Qy     390 GGCGCCGGGGAGGGCGGCGAGCACGGGCTGTTCGGGGGCCGGGCTCCCGGGGGGACTCGGG 449
Db   22407 GGCGCCGGGGAGGGCGGCGAGCACGGGCCGTTCGGGGGCCGGGCTCCCGGGGGGACTC 22348

Qy     450 GCACTCGGACGCGCAGGGCCACCTCACCTTCTCTTTTCCCTGGCCTCCCGGAGCAGCA 500
           ||||||||||| |||||||||||||||| |||||||| ||||||||||||||| |||||
Db   22347 GCACTCGGACGCGCAGGGCCACCTCACCTTCTCTTTTCCCTGGCCTTCCCGGAGCAGCA 22297
```

```
RESULT 9
HSU79264/c
HSU79264
LOCUS          HSU79264                1255 bp    mRNA       linear   PRI 28-NOV-2000
DEFINITION     Human clone 23814 mRNA sequence.
ACCESSION      U79264
VERSION        U79264.1  GI:1710223
KEYWORDS       .
SOURCE         human.
  ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 1255)
  AUTHORS      Andersson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.
  TITLE        A 'double adaptor' method for improved shotgun library construction
  JOURNAL      Anal. Biochem. 236 (1), 107-113 (1996)
  MEDLINE      96207227
REFERENCE      2  (bases 1 to 1255)
  AUTHORS      Yu,W., Andersson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,
               Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.
  TITLE        Large-scale concatenation cDNA sequencing
  JOURNAL      Genome Res. 7 (4), 353-358 (1997)
  MEDLINE      97264341
  PUBMED       9110174
REFERENCE      3  (bases 1 to 1255)
  AUTHORS      Yu,W. and Gibbs,R.A.
  TITLE        Direct Submission
  JOURNAL      Submitted (22-NOV-1996) Molecular and Human Genetics, Baylor
               College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
COMMENT        similar to human Zic protein mRNA sequence with GenBank Accession
               Number D76435.
FEATURES             Location/Qualifiers
     source          1. .1255
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="23814"
                     /sex="female"
                     /tissue_type="brain"
                     /clone_lib="Soares library 1NIB from IMAGE consortium"
                     /dev_stage="infant"
BASE COUNT       214 a    389 c    434 g    218 t
ORIGIN

Query Match 23.0%;  Score 114.8;  DB 9;  Length 1255;
Best Local Similarity 63.6%;  Pred. No. 2.9e-05;
Matches 220;  Conservative 0;  Mismatches 102;  Indels 24;  Gaps 2;

Qy     159 CGCCGCCCACATGGGAGCCTTCAAGCTCAAACCCGGCGCGCAGCGAGCGTGTCCCCGGCCA 218
           |||||  ||  ||||| ||||||||||||||||||  |||| ||| || | ||  ||||||
Db     857 CGCCGACGGCANGGGCGCCCTTCAAGCTCAAACCCCAGTTCGCACAGACGGGCTCCGGCCG 798

Qy     219 GAGCTCGGCGTTCACGTGCGGCAGGGCCCCGGGCCTACCCCGGCTCCGCCTCGCGGTGCCGC 278
           ||||| | || ||| || |||||||||| ||  ||| |          |||  ||||||
Db     797 CCAGACGGCGCCTTCACGTGCGCAGGGCGCCAGGCTAC------------GCGGCTGC 756

Qy     279 TGCGGCCGCAGCGCGTCGGGGGACTTCCTGTTCCGCCAGCGCGCCGCCGCACGTTGGCTCCTTCACTCTCTGGGCGCCCCTT 338
Db     755 TGCGGCCCCTTGGGCCATCACCACCCGGCGCCAGGCCA------------------GCGGCTGG 696

Qy     339 CAACTCACCCGGGACTTCCTGTTCCGCAGCGCGGCGCACGTTGGCTCCTTCACTCTCTGGGCGCCCCTT 398
Db     695 CAACTCCACGCGGGACTTTCTGTTCCGCAACCGGGTTTTGGCGACGCGGCGGGCGGCCT---------GCACCACGCGGCA 636

Qy     399 CGGCGGGCAGCACGGGGCGGCTGTTCGGGCCGGGCGGCGGGGCT---------GCACCACGCGGCA 452
Db     635 CACGCGACAGCACACGCCCTCTTTTGCTGCATCGGCGGTCGGGGGCGTTCGGGGGCCCACACGGCCA 576

Qy     453 CTCGGACGCGCAGGGCCACCTCTCTCTTCCCGGGGCCCTGCCAGAGCAG 498
           | ||||||||||||||||||||||||||||||||||||||||||||||
Db     575 CACGGACGCGCGGCGGGCCCACCTCCTTCTCCCGGGGCTTCCAGGAGCAG 530
```

```
RESULT 10
AC027060/c
LOCUS          AC027060                174559 bp    DNA        linear   HTG 19-APR-2000
DEFINITION     Homo sapiens chromosome 3 clone RP11-649A16 map 3, WORKING DRAFT
               SEQUENCE, 29 unordered pieces.
ACCESSION      AC027060
```

```
VERSION     AC027060.2  GI:7596863
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174559)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-649A16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174559)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 19, 2000 this sequence version replaced gi:7329421.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            -----------------------------------------------------------
                    Summary Statistics
            Sequencing Center: Whitehead Institute/ MIT Center for Genome Research
                    Center code: WIBR
                    Web site: http://www-seq.wi.mit.edu
                    Contact: sequence_submissions@genome.wi.mit.edu
            -----------------------------------------------------------
                    Project Information
            Center project name: L7415
            Center clone name: 649_A_16
            -----------------------------------------------------------
                    Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 157193 bases at least Q40
            Consensus quality: 166108 bases at least Q30
            Consensus quality: 166511 bases at least Q20
            Insert size: 170000; agarose-fp
            Insert size: 171759; sum-of-contigs
            Quality coverage: 4.1 in Q20 bases; agarose-fp
            Quality coverage: 4.1 in Q20 bases; sum-of-contigs
            -----------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1             1529: contig of 1529 bp in length
            *     1530   1629: gap of 100 bp
            *     1630   2793: contig of 1164 bp in length
            *     2794   2893: gap of 100 bp
            *     2894   4504: contig of 1611 bp in length
            *     4505   4604: gap of 100 bp
```

```
FEATURES             Location/Qualifiers
     source          1..174559
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-649A16"
                     /map="3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1529
                     /note="assembly_fragment"
     misc_feature    1630..2793
                     /note="assembly_fragment"
     misc_feature    2894..4504
                     /note="assembly_fragment"
     misc_feature    4605..6652
                     /note="assembly_fragment"
     misc_feature    6753..8434
                     /note="assembly_fragment"
     misc_feature    8535..10636
                     /note="assembly_fragment"
     misc_feature    10737..12486
                     /note="assembly_fragment"
```

```
            *     4605             6652: contig of 2048 bp in length
            *     6753   6752: gap of 100 bp
            *     6753   8434: contig of 1682 bp in length
            *     8435   8534: gap of 100 bp
            *     8535   10636: contig of 2102 bp in length
            *    10637  10736: gap of 100 bp
            *    10737  12486: contig of 1750 bp in length
            *    12487  12586: gap of 100 bp
            *    12587  15295: contig of 2709 bp in length
            *    15296  15395: gap of 100 bp
            *    15396  17556: contig of 2161 bp in length
            *    17557  17656: gap of 100 bp
            *    17657  21289: contig of 3633 bp in length
            *    21290  21389: gap of 100 bp
            *    21390  25103: contig of 3714 bp in length
            *    25104  25203: gap of 100 bp
            *    25204  28155: contig of 2952 bp in length
            *    28156  28255: gap of 100 bp
            *    28256  30966: contig of 2711 bp in length
            *    30967  31066: gap of 100 bp
            *    31067  33640: contig of 2574 bp in length
            *    33641  33740: gap of 100 bp
            *    33741  38356: contig of 4616 bp in length
            *    38357  38456: gap of 100 bp
            *    38457  43253: contig of 4797 bp in length
            *    43254  43353: gap of 100 bp
            *    43354  49544: contig of 6191 bp in length
            *    49545  49644: gap of 100 bp
            *    49645  53734: contig of 4090 bp in length
            *    53735  53834: gap of 100 bp
            *    53835  58998: contig of 5164 bp in length
            *    58999  59098: gap of 100 bp
            *    59099  65338: contig of 6240 bp in length
            *    65339  65438: gap of 100 bp
            *    65439  71814: contig of 6376 bp in length
            *    71815  71914: gap of 100 bp
            *    71915  76662: contig of 4748 bp in length
            *    76663  76762: gap of 100 bp
            *    76763  83989: contig of 7227 bp in length
            *    83990  84089: gap of 100 bp
            *    84090  94338: contig of 10249 bp in length
            *    94339  94438: gap of 100 bp
            *    94439  104432: contig of 9994 bp in length
            *   104433  104532: gap of 100 bp
            *   104533  118453: contig of 13921 bp in length
            *   118454  118553: gap of 100 bp
            *   118554  131437: contig of 12784 bp in length
            *   131438  131537: gap of 100 bp
            *   131538  152159: contig of 20722 bp in length
            *   152160  152259: gap of 100 bp
            *   152260  174555: contig of 22300 bp in length.
```

```
misc_feature    12587..15295
                /note="assembly_fragment"
misc_feature    15396..17556
                /note="assembly_fragment"
misc_feature    17657..21289
                /note="assembly_fragment"
misc_feature    21390..25103
                /note="assembly_fragment"
misc_feature    25204..28155
                /note="assembly_fragment"
misc_feature    28256..30966
                /note="assembly_fragment"
misc_feature    31067..33640
                /note="assembly_fragment"
                /clone_end:T7
                vector_side:left"
misc_feature    33741..38356
                /note="assembly_fragment"
misc_feature    38457..43253
                /note="assembly_fragment"
misc_feature    43354..49544
                /note="assembly_fragment"
misc_feature    49645..53734
                /note="assembly_fragment"
misc_feature    53835..58998
                /note="assembly_fragment"
                /clone_end:SP6
                vector_side:right"
misc_feature    5099..65338
                /note="assembly_fragment"
misc_feature    65439..71814
                /note="assembly_fragment"
misc_feature    71915..76662
                /note="assembly_fragment"
misc_feature    76763..83989
                /note="assembly_fragment"
misc_feature    84090..94338
                /note="assembly_fragment"
misc_feature    94439..104432
                /note="assembly_fragment"
misc_feature    104533..11453
                /note="assembly_fragment"
misc_feature    11854..131337
                /note="assembly_fragment"
misc_feature    131438..152159
                /note="assembly_fragment"
misc_feature    152260..174559
                /note="assembly_fragment"
BASE COUNT      51760 a  34803 c  33897 g  51293 t  2806 others
ORIGIN

Query Match   23.0%;  Score 114.8;  DB 2;  Length 174559;
Best Local Similarity  63.6%;  Pred. No. 6.1e-06;
Matches 220;  Conservative  0;  Mismatches 102;  Indels 24;  Gaps  2;

Qy    159 CGCCGCGCACATGGGAGGCCTTCAAGCTCAACCCGGGCCGCGACGAGCTGGCTTCCGGGGCCA 218
          |||||       ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db 155691 CGCCGACGGCGTTCAGGTGGCGGCCTTCAAGCTCAACCCCAGTTCGCAGAGCTGGCTTCGGCCG 155632
Qy    219 GAGCTCGGCGCTTCACGTCGCAGGGCGCCCGGGCCTCCCGCGTGCCGGCTGCCGC 278
          ||||||||||||||||||||| | ||| |||
Db 155631 CCAGACAGGCCTTCACGTCGCAGGGCGCCCAGGCTAC------GCGGCTGC 155590
Qy    279 TGCGGCGCCGCCGCAGCGCTCGGGCGCCCCACCGCGCGCCGCGCCCCTT 338
          ||||||||       |
Db 155589 TGCCGGCCCTGGGCCATCACCACCGGCTCCTATTCCAGCGGCCCTT 155530
Qy    339 CAACTCCACCCGGAGACTTCCTCGTTCCGAGGACTTCGGGCCGGG 398
          ||||||||||||| | ||| |||| | |||||  ||||
Db 155529 CAACTCCACCACCGGGACTTTCTGTTCCGCAACCGGCCGCGGGCCA 155470
Qy    399 CGGCGGCGCAGCACGGGCGGCTGTTCGGGCCGGGGGCCT 452
          |||| |||||||||||||||||||||||||||| || |
Db 155529 CAACTCCACCACCGGGACTTTCTGTTCCGCAACCGGCCGCGGGCCA 155470
```

```
Db 155469 CAGCGCACAGCAGCAGCGCTCTTTGCTGCATCGGCGGCGGCTTGCCAGAGCAG 498
          |||||||||||||||||||||||||| ||||| ||||||||||| || ||||
Db 155409 CACGGACGCCGCGGCGGCCACCTCTCCGGGGCCTTCACGAGCAG 155364

RESULT 11
LOCUS           AR103244                     3138 bp     DNA      linear   PAT 14-FEB-2001
DEFINITION      Sequence 5 from patent US 6087168.
ACCESSION       AR103244
VERSION         AR103244.1  GI:12814832
KEYWORDS
SOURCE          Unknown.
  ORGANISM      Unknown.
                Unclassified.
REFERENCE       1  (bases 1 to 3138)
  AUTHORS       Levesque,M.F. and Neuman,T.
  TITLE         Conversion of non-neuronal cells into neurons: transdifferentiation
                of epidermal cells
  JOURNAL       Patent: US 6087168-A 5 11-JUL-2000;
FEATURES             Location/Qualifiers
     source          1..3138
                     /organism="unknown"
BASE COUNT      701 a   898 c   840 g   699 t
ORIGIN

Query Match   22.6%;  Score 113.2;  DB 6;  Length 3138;
Best Local Similarity  63.3%;  Pred. No. 3.4e-05;
Matches 219;  Conservative  0;  Mismatches 103;  Indels 24;  Gaps  2;

Qy    159 CGCCGCGCACATGGGAGGCCTTCAAGCTCAACCCGGGCCGCGACGAGCTGGCTTCCGGGGCCA 218
          |||||       ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    900 CGCCGACGGCGTTCAGGTGGCGGCCTTCAAGCTCAACCCCAGTTCGCAGAGCTGGCTTCGGCCG 959
Qy    219 GAGCTCGGCGCTTCACGTCGCAGGGCGCCCGGGCCTCCCGCGTGCCGGCTGCCGC 278
          ||||||||||||||||||||| | ||| |||
Db    960 CCAGACAGGCCTTCACGTCGCAGGGCGCCCAGGCTAC------GCGGCTGC 1001
Qy    279 TGCGGCGCCGCCGCAGCGCTCGGGCGCCCCACCGCGCGCCGCGCCCCTT 338
          ||||||||       |
Db   1002 TGCCGGCCCTGGGCCATCACCACCGGCTCCTATTCCAGCGGCCCTT 1061
Qy    339 CAACTCCACCCGGAGACTTCCTCGTTCCGAGGACTTCGGGCCGGG 398
          ||||||||||||| | ||| |||| | |||||  ||||
Db   1062 CAACTCCACCACCGGGACTTTCTGTTCCGCAACCGGCCGCGGGCCA 1121
Qy    399 CGGCGGCGCAGCACGGGCGGCTGTTCGGGCCGGGGGCCT 452
          |||| |||||||||||||||||||||||||||| || |
Db   1122 CAGCGCACAGCAGCAGCGCTCTTTGCTGCATCGGCGGCGGCTTGCCAGAGCAG 1181
Qy    453 CTCGGACGCGCAGGGCCACCTCTCCGGGGCCTGCCAGAGCAG 498
          ||||||||       |||||||||||||||||||| |||||
Db   1182 CACGGACGCCGCGGCGGCCACCTCTCCGGGGCCTTCACGAGCAG 1227

RESULT 12
LOCUS           BD000107                     3138 bp     DNA      linear   PAT 31-JAN-2002
DEFINITION      Differentiation and transformation of transfected epithelial basal
                cell into neural progenitor cell, neuron cell and/or glia cell.
ACCESSION       BD000107
VERSION         BD000107.1  GI:18623186
KEYWORDS        JP 2002295987-A/5.
SOURCE          Homo sapiens
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 3138)
  AUTHORS       Rebesuku,M.F. and Newman,T.
  TITLE         Differentiation and transformation of transfected epithelial basal
                cell into neural progenitor cell, neuron cell and/or glia cell
```

JOURNAL   Patent: JP 200295987-A 5 24-OCT-2000;
COMMENT   CEDARS SINAI MEDICAL CENTER
OS   Homo sapiens (human)
PN   JP 200295987-A/5
PD   24-OCT-2000
PF   20-JAN-2000 JP 2000048291
PR   20-JAN-1999 US   234332
PI   MICHEL F REBESHKU, THOMAS NEWMAN
PC   C12N15/02,A61K35/30,A61P25/28,A61P43/00,C12N15/00,C12N5/10,C12Q1/02, PC
PC   G01N33/50//C07K14/52,C12P21/02,C12N15/00,C12N5/00 CC
PC   G01N33/15;

FEATURES   Location/Qualifiers
source   1..3138
           /organism="Homo sapiens"
           /db_xref="taxon:9606"

FT   Key   Location/Qualifiers
FT   gene   (0)..(0).
      G01N33/50//C07K14/52,C12P21/02,C12N15/00,C12N5/00 CC

BASE COUNT   701 a   898 c   840 g   699 t
ORIGIN

Query Match   22.6%;   Score 113.2;   DB 6;   Length 3138;
Best Local Similarity   63.3%;   Pred. No. 3.4e-05;
Matches 219;   Conservative   0;   Mismatches 103;   Indels 24;   Gaps 2;

Qy   159 CGCCGGCCACATGGGAGGCCTTCAAGCTCAACACCCGGCGCGCGACGAGCTGTCCCGGGCCA 218
          ||| |||||||||||| |||||||||||||||||| | |||||||||||||||||||||||
Db   900 CGCCGACGGCCATGGGCGCGCCTTCAAGCTCAACACCCCAGTTCGCACGAGCTGGCTTCGGCCG 959

Qy   219 GAGCTCGGCGTTCACGTCGCAGGGCCCGGGCCTCACCCCGGCGTCCCGTGCCGCTGCCGCC 278
          ||||||||||||||||||||| ||| ||||||||||||| |||| |||||||||| |||
Db   960 CCAGACAGCCTTCACGTCGCAGGGCCTAC-------------GCGGCTGC 1001

Qy   279 TGCGGGCCGCGACGCCTCGGGGCCCCCACGCCGCGCGACGTTGGCTCCTACTCTGGGCGCC 338
          ||||||| ||||||| |||||||||| |||||||||||| | | |||||| |||||
Db   1002 TGCGGGCCCTGGGCGCCATCACCATCACCCGGCCGCGCGACGTCGGCTCCTATTCCAGCGCGCAGCCTT 1061

Qy   339 CAACTCACCCCGGGACTTCCTGTTCCGGACCGGCGGCGGACTTGCGGCCGCCGCCGCCGG 398
          ||||||||||||||||||| |||||||||||||||||| |||||||||| |||||||
Db   1062 CAACTCCACGGCGGGGACTTTCGTTCCCAACCACGGGGTTTTGGCGACCGCGGCGGCCGCGCC 1121

Qy   399 CGGCGGGGCAGCACGGGGCTGTGTTCGGGCCGCGGGGCGGCGGCCCT-----GCACCACGCGCA 452
          ||| |||||| || ||| ||||||| ||||||||||||||||| | |||| |||||||
Db   1122 CAGCGCACGACAGCCTTCACGTCGCACGGCGCCGCCGCCAGGCTAC-------------GCGGCTGC 1181

Qy   453 CTCGGACGCGCAGGGGCCCACCCTCTTCTTTGCGATCGGGCCGGGGCGGGGCCCGGG 498
          | |||||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db   1182 CACGGACGCGCGCGGGCGCCACCCTCTTCTTCCCGGGGCTTCACGAGCAG 1227

RESULT 13
LOCUS   HUM2ICP   3138 bp   mRNA   linear
DEFINITION Homo sapiens mRNA for zic protein, complete cds.
ACCESSION D76435
VERSION   D76435.1 GI:1208428
KEYWORDS   zic protein; zic.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens cerebellum cDNA to mRNA.

REFERENCE 1 (bases 1 to 3138)
AUTHORS   Yokota,N.
JOURNAL   Direct Submission
           Submitted (16-OCT-1995) Naoki Yokota, Institute of Physical and
           Chemical Research (RIKEN), Molecular Neurobiology Laboratory; 3-1-1
           Koyadai, Tsukuba, Ibaraki 305, Japan (Tel:0298-36-9170,
           Fax:0298-36-9040)
REFERENCE 2 (bases 1 to 3138)
AUTHORS   Yokota,N.
JOURNAL   Unpublished (1996)
REFERENCE 3 (sites)
AUTHORS   Yokota,N., Aruga,J., Takai,S., Yamada,K., Hamazaki,M., Iwase,T.,
           Suginura,H. and Mikoshiba,K.

TITLE   Predominant expression of human zic in cerebellar granule cell
           lineage and medulloblastoma
JOURNAL   Cancer Res. 56 (2), 377-383 (1996)
MEDLINE   96134385

FEATURES   Location/Qualifiers
source   1..3138
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="3"
           /map="3q24"
           /tissue_type="cerebellum"
gene   781..2124
           /gene="zic"
CDS   781..2124
           /gene="zic"
           /codon_start=1
           /product="zic protein"
           /protein_id="BAA11179.1"
           /db_xref="GI:1208429"
           /translation="MLLDAGPQYPAIGVTFGASRHHSAGDVAERDVGLGINPFADGM
           GAFKLNPSSHELASAGQTAPTSOAPGYAAAAALGHHHHPGHVGSYSSAAPNSTRDFLF
           RNRGFGDAAAAASAQHSLFAASAGGFGFGGHGHHTDAAGHLLFPGLHEQAAGHASPNVVN
           GQMRLGFSGDMYPRPEQYGQVTSPRSEHYAAPQLHGYGPMNVNMAAHHQAGAFRYMR
           QPIKQELICKWIEPEQLANPKKSCNKTFSTMHELVTHVTVEHVGGPEQSNHICEWEEC
           PR£GKPFKAKYKLVNHIRVHTGEKPFPCPFPGCGKVFAARSENLKIHKRTHTGEKPFKC
           EFEGCDRRFANSSDRKKHMHVHTSDKPYLCKMCDKSYTHPSSVRKHMKVHESSSGQSQ
           PSPAASSGYESSTPPTIVSPSTDNPTTSSLSPSSSAVHHFAGHSALSSNPNEWYY"

BASE COUNT   701 a   898 c   840 g   699 t
ORIGIN   Chromosome 3q24.

Query Match   22.6%;   Score 113.2;   DB 9;   Length 3138;
Best Local Similarity   63.3%;   Pred. No. 3.4e-05;
Matches 219;   Conservative   0;   Mismatches 103;   Indels 24;   Gaps 2;

Qy   159 CGCCGGCCACATGGGAGGCCTTCAAGCTCAACACCCGGCGCGCGACGAGCTGTCCCGGGCCA 218
          ||| |||||||||||| |||||||||||||||||| | |||||||||||||||||||||
Db   900 CGCCGACGGCCATGGGCGCGCCTTCAAGCTCAACACCCCAGTTCGCACGAGCTGGCTTCGGCCG 959

Qy   219 GAGCTCGGCGTTCACGTCGCAGGGCCCGGGCCTCACCCCGGCGTCCCGTGCCGCTGCCGCC 278
          ||||||||||||||||||||| ||| ||||||||||||| |||| ||||||||| |||
Db   960 CCAGACAGCCTTCACGTCGCAGGGCCGCCGTAC-------------GCGGCTGC 1001

Qy   279 TGCGGGCCGCGACGCCTCGGGGCCCCCACGCCGCGCGACGTTGGCTCCTACTCTGGGCGCC 338
          ||||||| ||||||| |||||||||| |||||||||| | | |||||| |||||
Db   1002 TGCGGGCCCTGGGCGCCATCACCATCACCCGGCCGCGCGACGTCGGCTCCTATTCCAGCGCGCAGCCTT 1061

Qy   339 CAACTCACCCCGGGACTTCCTGTTCCGGACCGGCGGCGGACTTGCGGCCGCCGCCGCCGG 398
          ||||||||||||||||||| |||||||||||||||||| |||||||||| |||||||
Db   1062 CAGCGCCACGACAGCCTTCACGTCGCAGGGCGCCGCCGCCAGGCTAC-----GCGGCTGC 1121

Qy   399 CGGCGGGGCAGCACGGGGCTGTGTTCGGGCCGCGGGGCGGCGGCCCT-----GCACCACGCGCA 452
          ||| |||||| || ||| ||||||| |||||||||||||||| | |||| |||||||
Db   1122 CAGCGCCACACAGCCTCTTTGCTCGATCACCACCCGGCCGCGCGACGTTGGCTCCTACTCTGGGGCCCCTT 1181

Qy   453 CTCGGACGCGCAGGGGCCCACCCTCTTCTTTGCGATCGGGCCGGGGCGGGGCCCGGG 498
          | |||||| ||||||||||||||||||||||||| ||||||||||||||||||||
Db   1182 CACGGACGCGCGCGGGCGCCACCCTCTTCTTCCCGGGGCTTCACGAGCAG 1227

RESULT 14
LOCUS   AC095557/c   59150 bp   DNA   linear   HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-8D4, *** SEQUENCING IN PROGRESS ***,
            34 unordered pieces.
ACCESSION AC095557
VERSION   AC095557.2 GI:17942085
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
ORGANISM   Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 59150)

```
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
          Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
          Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
          Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
          Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
          Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
          Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
          Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
          Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
          Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
          Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
          Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
          Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
          Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
          Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
          Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
          Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
          Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
          Joudah,S., Karlson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
          Kovar,C., Kratovic,J., Kureshi,N., Landry,N., Leal,B., Lewis,L.C.,
          Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
          Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
          Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
          Martinez,E., Massey,E., Mawhiney,E., McLeod,M., Meador,M.,
          Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
          Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,N.,
          Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
          Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
          Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
          Puiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
          Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
          Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,S.,
          Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
          Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
          Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
          Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
          Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
          Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 2  (bases 1 to 59150)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Dec 20, 2001 this sequence version replaced gi:15627177.
          --------------- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
          --------------- Project Information
             Center project name: GCRP
             Center clone name: CH230-8D4
          --------------- Summary Statistics
             Assembly program: Phrap; version 0.990329First call to
          findPhrapList*
             Consensus quality: 53080 bases at least Q40
             Consensus quality: 57822 bases at least Q30
             Consensus quality: 61489 bases at least Q20
             Estimated insert size: 40280; sum-of-contigs estimation
             Quality coverage: 0x in Q20 bases; agarose-fp estimation
             Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
          ---------------
          * NOTE: Estimated insert size may differ from sequence length
          *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *    consists of 34 contigs. The true order of the pieces
          *    is not known and their order in this sequence record is
          *    arbitrary. Gaps between the contigs are represented as
          *    runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *     1    3299: contig of 3299 bp in length
          *  3300    3399: gap of unknown length
          *  3400    5449: contig of 2050 bp in length
          *  5450    5549: gap of unknown length
          *  5550    7088: contig of 1539 bp in length
          *  7089    7188: gap of unknown length
          *  7189    9405: contig of 2217 bp in length
          *  9406    9505: gap of unknown length
          *  9506   12651: contig of 3146 bp in length
          * 12652   12751: gap of unknown length
          * 12752   15134: contig of 2383 bp in length
          * 15135   15234: gap of unknown length
          * 15235   17420: contig of 2186 bp in length
          * 17421   17520: gap of unknown length
          * 17521   20217: contig of 2697 bp in length
          * 20218   20317: gap of unknown length
          * 20318   22134: contig of 1817 bp in length
          * 22135   22234: gap of unknown length
          * 22235   23804: contig of 1570 bp in length
          * 23805   23904: gap of unknown length
          * 23905   25367: contig of 1463 bp in length
          * 25368   25467: gap of unknown length
          * 25468   27472: contig of 2005 bp in length
          * 27473   27572: gap of unknown length
          * 27573   29410: contig of 1838 bp in length
          * 29411   29510: gap of unknown length
          * 29511   30798: contig of 1288 bp in length
          * 30799   30898: gap of unknown length
          * 30899   32015: contig of 1117 bp in length
          * 32016   32115: gap of unknown length
          * 32116   33224: contig of 1109 bp in length
          * 33225   33324: gap of unknown length
          * 33325   34615: contig of 1291 bp in length
          * 34616   34715: gap of unknown length
          * 34716   36067: contig of 1352 bp in length
          * 36068   36167: gap of unknown length
          * 36168   37648: contig of 1481 bp in length
          * 37649   37748: gap of unknown length
          * 37749   38981: contig of 1233 bp in length
          * 38982   39081: gap of unknown length
          * 39082   40238: contig of 1157 bp in length
          * 40239   40338: gap of unknown length
          * 40339   41420: contig of 1082 bp in length
          * 41421   41520: gap of unknown length
          * 41521   42879: contig of 1359 bp in length
          * 42880   42979: gap of unknown length
          * 42980   44278: contig of 1299 bp in length
          * 44279   44378: gap of unknown length
          * 44379   46353: contig of 1975 bp in length
          * 46354   46453: gap of unknown length
          * 46454   47698: contig of 1245 bp in length
          * 47699   47798: gap of unknown length
          * 47799   49472: contig of 1674 bp in length
          * 49473   49572: gap of unknown length
          * 49573   51018: contig of 1446 bp in length
          * 51019   51118: gap of unknown length
          * 51119   52548: contig of 1430 bp in length
          * 52549   52648: gap of unknown length
          * 52649   53839: contig of 1191 bp in length
          * 53840   53939: gap of unknown length
          * 53940   55252: contig of 1313 bp in length
          * 55253   55352: gap of unknown length
          * 55353   56359: contig of 1007 bp in length
          * 56360   56459: gap of unknown length
          * 56460   57593: contig of 1134 bp in length
          * 57594   57693: gap of unknown length
          * 57694   59150: contig of 1457 bp in length
FEATURES             Location/Qualifiers
     source          1..59150
                     /organism="Rattus norvegicus"
```

FEATURES   Location/Qualifiers
     source   1..1344
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
     gene   1..1344
                     /gene="zic"
     CDS   1..1344
                     /gene="zic"
                     /codon_start=1
                     /product="zic protein"
                     /protein_id="AAF34656.1"
                     /db_xref="GI:6979926"
                     /translation="MLLDAGPQYPAIGVTTFGASRHHSAGDVAERDVGLGINPFADGM
                     GAFKLNPSSHELASAGQTAFTSQAPGYAAAAALGHHHHPGHVGYSYSSAAFNSTRDFLF
                     RNRGFGDAAAAASAQHSLFAASAGFGPGHGHTDAAGHLLFSGLHEQAAGHASPNVVN"

RESULT 15
LOCUS   AF221839   1344 bp   mRNA   linear   ROD 16-FEB-2000
DEFINITION Rattus norvegicus zic protein (zic) mRNA, complete cds.
ACCESSION   AF221839
VERSION   AF221839.1 GI:6979925
KEYWORDS   .
SOURCE   Norway rat.
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 1344)
  AUTHORS   Kang,J.S., Cai,Y.Y., Liu,H.I. and Li,R.X.
  TITLE   A novel zic gene from SD rat
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1344)
  AUTHORS   Kang,J.S., Cai,Y.Y., Liu,H.I. and Li,R.X.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-JAN-2000) Shanghai Reseach Center of Biotechnology,
            Chinese Academy of Science, 500# CaoBao Road, Shanghai, Shanghai
            200233, China

BASE COUNT   15992 a   11613 c   11687 g   16500 t   3358 others
ORIGIN
          /db_xref="taxon:10116"
          /clone="CH230-8D4"

Query Match   21.4%; Score 106.8; DB 2; Length 59150;
Best Local Similarity   62.1%; Pred. No. 7.6e-05;
Matches 215; Conservative 0; Mismatches 107; Indels 24; Gaps 2;

Qy 159 CGCCCGGCCACATGGGAGGAGCCTTCAAGCTCAACACCCGGCGCGGACGAGCTGTCCCCGGGCCA 218
        ||||| |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 6934 CGCCGGACGGCATGGGCGGGGCCTTCAAAGCTCAACCCCAGTTGCCACGAACTGGCCTCGGCCGG 6875

Qy 219 GAGCTCGGGCGTTCACGTCGCAGGGGCCCCGGGGGCCTACCCCCGGCTGCCGGCGGGTGCCGC 278
        ||| ||||||||||||||||||||||||| || || |||||||||||||||| |||||||
Db 6874 CCAGACAGCCTTCACGTCGCAGGCCTCCGGGGCTAC------------------GCCGCTGC 6833

Qy 279 TGCCGGCTGCCGCAGCGCGTCGGGGCCCCCACCCGCGCGCACGTTGGCTGGCCGCCCCTT 338
        |||||| |||| ||||||||||||||||||||||||||||||| ||
Db 6832 CGCGGCCCTGGGCACCATCACCACCCGCGCACGTCGGCTCCGGGCTAC---------GCCGCTGC 6773

Qy 339 CAACTCCACCCGGGACTTCCTGTTCCGCAGCGCCGCTTCCGGGGACTTCGGCCCGGG 398
        ||||| ||||||||||||||||||||||||||||| ||| | |||||||| |
Db 6772 CAATTCTACCCGGGACTTTCTGTTCCGCAACCGGGGCTTCGGCGACGCGGCGGCAGC 6713

Qy 399 CGGCGGGCAGCGACGGGGCTGTTCGGGGGCGGCGGGCGGCCGCCT------GCACCACGGCA 452
        ||| ||| | |||||||||||||||||||||||||||||||||| || |||||| ||
Db 6712 TAGCGCCAGCACACAGTCTCTTCCGGGGGCCCTTCGGGGGGGCCCACACGGCA 6653

Qy 453 CTCGGACGCGCGCAGGGCCCACCTCCTCCTTCTTTTTTCTGGGGCTTCACGAGACGAG 498
        |||| ||||||||||||||||| |||||||||||||||||||||||||||
Db 6652 TACGGACGCGCGCGGGGCCACCTTCTTTTTTCTGGGGCTTCACGAGCAG 6607

GQMRLGFSGDMYPRPEQYGQVTISPRSEHYAAPQLHGYGPMNVNMAAHHGAGAFFRYMR
QPIKQELICKWIEPEQLANPKKSCNKTFSTMHELVTHVTVEHVGGPEQSNHICFWEEC
PREGKPFKAKYKLVNHIRVHTGEKPFPCPFPGGCKVFARSENLKIHKRTHTGEKPFKC
EFEGCDRRFANSSDRKKHMVHTSDKPYLCKMCDKSYTHPSSLRKHMKVHESSSQGSQ
PSPAASSGYESSTPPTIVSPTTDNPTTSSMSPSSSAVHHTAGHSALSSNFNEWYV"

BASE COUNT   278 a   478 c   359 g   229 t
ORIGIN

Query Match   21.0%; Score 105.2; DB 10; Length 1344;
Best Local Similarity   61.8%; Pred. No. 0.00039;
Matches 214; Conservative 0; Mismatches 108; Indels 24; Gaps 2;

Qy 159 CGCCCGGCCACATGGGAGGAGCCTTCAAGCTCAACACCCGGCGCGGACGAGCTGTCCCCGGGCCA 218
        ||||| |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 120 CGCCGGACGGCATGGGCGGGGCCTTCAAAGCTCAACCCCAGTTGCCACGAACTGGCCTCGGCCGG 179

Qy 219 GAGCTCGGGCGTTCACGTCGCAGGGGCCCCGGGGGCCTACCCCCGGCTGCCGGCGGGTGCCGC 278
        ||| |||||||||||||||||||||||| || || |||||||||||||||| |||||||
Db 180 CCAGACAGCCTTCACGTCGCAGGCCTCCGGGGCTAC------------------GCCGCTGC 221

Qy 279 TGCCGGCGGCCGCAGCGCGTCGGGGCCCCCACCCGCGCGCACGTTGGCTCCTGACTTGGCGCCCCTT 338
        |||||| |||| ||||||||||||||||||||||||||||||||||| ||
Db 222 CGCGGCCCTGGGCACCATCACCACCCGCGCACGTCGGCTCCGGGCTAC---------GCCGCTGC 281

Qy 339 CAACTCCACCCGGGACTTCCTGTTCCGCAGCGCCGCTTCCGGGGACTTCGGCCCGGG 398
        ||||| ||||||||||||||||||||||||||||| ||| | |||||||| |
Db 282 CAATTCTACCCGGGACTTTCTGTTCCGCAACCGGGGCTTCGGCGACGCGGCGGCAGC 341

Qy 399 CGGCGGGCAGCGACGGGGCTGTTCG------GGCGGCGGGCGGCCGCCTGCACCACGGCA 452
        ||| ||| | |||||||||||||| ||||||||||||| ||| |||||| ||
Db 342 TAGCGCCAGCACACAGTCTCTTCCGCTGCTTCGGGGGGGCCCACACGGCA 401

Qy 453 CTCGGACGCGCGCAGGGCCCACCTCCTCCTTCTTTTTTCTGGGGCTTCACGAGACGAG 498
        |||| ||||||||||||||||| |||||||||||||||||||||||||||
Db 402 TACGGACGCGCGCGGGGCCACCTTCTTTTTTCTGGGGCTTCACGAGCAG 447

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                 (without alignments)
                 14616.225 Million cell updates/sec

Title: US-09-489-101A-6_COPY_1_500
Perfect score: 500
Sequence: 1 ggaattccgggcgcggttgt.........ggtggctccgcagccgccggc 500

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*

#							
1:	gb_ba:*
2:	gb_htg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vi:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*
29:	em_vi:*
30:	em_fun:*
31:	em_ba:*
32:	em_htg_hum:*
33:	em_htg_inv:*
44:	em_htg_other:*
45:	em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		500	100.0	1322	6	AX201585	AX201585 Sequence
2		500	100.0	1322	9	HSU28368	U28368 Human Id-re
3		477.4	95.5	2389	9	BC014941	BC014941 Homo sapi
4	c	477.4	95.5	185371	9	HS625H18	AL022726 Human DNA
5	c	316.2	63.2	1017	9	HSU16153	U16153 Human Id-4H
6	c	288.2	57.6	1473	9	AF030295	AF030295 Homo sapi
7	c	189	37.8	1473	6	AX323847	AX323847 Sequence
8	c	177.4	35.5	176835	9	AC099470	AC099470 Rattus no
9	c	171.8	34.4	4701	10	MMHLH1972	MMHLH1972 Homo
10	c	170.2	34.0	4503	10	AF077859	AF077859 Mus muscu
11		158.4	31.7	1147	6	HSID4	Y07958 H.sapiens m
12		149.6	29.9	1473	6	AX323848	AX323848 Sequence
13		133.2	26.6	1659	10	MMID4HLH	X75018 M.musculus
14		116.8	23.4	1309	6	ARI03268	AR103268 Sequence
15	c	66.4	13.3	195286	2	AC097157	AC097157 Rattus no
16		65.2	13.0	161307	2	AC084058	AC084058 Rattus no
17		64.2	12.8	181191	2	AC080138	AC080138 Mus muscu
18		63.4	12.7	1055	10	MMID4	X96729 M.musculus
19	c	62.2	12.4	144505	2	AL662884	AL662884 Homo sapi
20		61.2	12.2	90005	2	AC106630	AC106630 Rattus no
21		61	12.2	63629	2	AC091187	AC091187 Homo sapi
22		59.4	11.9	94526	2	AC096092	AC096092 Rattus no
23		59.4	11.9	101624	2	AC106598	AC106598 Rattus no
24	c	59.4	11.9	179825	2	AC009672	AC009672 Homo sapl
25		58.8	11.8	183584	2	AC095195	AC095195 Rattus no
26	c	58.6	11.7	52664	2	AC097284	AC097284 Rattus no
27	c	58.4	11.7	172506	2	AC106467	AC106467 Rattus no
28		58	11.6	145427	2	AC095796	AC095796 Rattus no
29	c	57.4	11.5	166036	2	AC094985	AC094985 Rattus no
30	c	57.4	11.5	120316	2	AC025984	AC025984 Homo sapi
31		57.2	11.4	200066	2	AC016520	AC016520 Homo sapi
32	c	57.2	11.4	182589	2	AC009954	AC009954 Homo sapi
33	c	56.8	11.4	125020	9	AF429315	AF429315 Homo sapi
34	c	56.4	11.3	205914	2	AC027682	AC027682 Homo sapi
35		56.2	11.2	86621	2	AC093933	AC093933 Rattus no
36	c	56	11.2	145427	2	AC095796	AC095796 Rattus no
37		55.8	11.2	217412	2	AC024400	AC024400 Homo sapi
38	c	55.2	11.0	90922	2	AC094187	AC094187 Rattus no
39	c	55.2	11.0	147524	2	AC069201	AC069201 Homo sapi
40	c	55	11.0	57121	2	AC084255	AC084255 Homo sapi
41		55	11.0	91532	2	AC108628	AC108628 Homo sapi
42		55	11.0	161307	2	AC084058	AC084058 Homo sapi
43		54.4	10.9	60583	2	AC036123	AC036123 Homo sapi
44		54.4	10.9	153170	2	AC103491	AC103491 Rattus no
45		54.2	10.8	144306	2	AC099248	AC099248 Rattus no

ALIGNMENTS

RESULT 1
AX201585
LOCUS   AX201585   1322 bp   DNA   linear   PAT 30-AUG-2001
DEFINITION Sequence 6 from Patent WO0153349.
ACCESSION   AX201585
VERSION   AX201585.1 GI:1539l434
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1322)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE   Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 6 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES   Location/Qualifiers
     source   1..1322
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

```
BASE COUNT      267 a      376 c      417 g      262 t
ORIGIN

Query Match                100.0%;  Score 500;  DB 6;  Length 1322;
Best Local Similarity      100.0%;  Pred. No. 9.3e-72;
Matches    500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGAATTCCGGGCGCCGGTTGTGAGTAGTACCGGGAGTGGGGTGATCCCGGGGCTAGGGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAATTCCGGGCGCCGGTTGTGAGTAGTACCGGGAGTGGGGTGATCCCGGGGCTAGGGGAGC 60

Qy     61 GCGGGCGCCCGATCGGGCGGCTTAGTGCGGAGCTCCGAAGGAGGACACCCGGGGTGGG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGGGCGCCCGATCGGGCGGCTTAGTGCGGAGCTCCGAAGGAGGACACCCGGGGTGGG 120

Qy    121 CTACTTTTTCTTCCGGTGCTTTTGCTTTTTTGGGTCGGGCTTGAGTGTCGGTCGCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTACTTTTTCTTCCGGTGCTTTTGCTTTTTTGGGTCGGGCTTGAGTGTCGGTCGCCC 180

Qy    181 ACTGACAAAAGATTCCCTCGTAAAACCCAGAGCGACACCCTCCCGTCAATTGTTGGGCTCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACTGACAAAAGATTCCCTCGTAAAACCCAGAGCGACACCCTCCCGTCAATTGTTGGGCTCGG 240

Qy    241 GAGTGTCGCGGTGCCCCGAGCGGGCGGGAGGCAAAGGAGGAGCCGGCCCGCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGTGTCGCGGTGCCCCGAGCGGGCGGGAGGCAAAGGAGGAGCCGGCCCGCG 300

Qy    301 GACGGGGGCCCGGAGCTTGGCCTCGCCTCGCGGCGGCGGGTTGCGCTCGCGGTAGAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGGGGCCCGGAGCTTGGCCTCGCCTCGCGGCGGCGGGTTGCGCTCGCGGTAGAG 360

Qy    361 CGCAGGGGGCGCGCGACTGAAGCGGGTGAGCCCCTCCGGCCCAAGGCGCCG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGCAGGGGGCGCGCGACTGAAGCGGGTGAGCCCCTCCGGCCCAAGGCGCCG 420

Qy    421 TCGGGCTGCGCGGCGGGGAGCTGGCCGCTGCGCCGCCACGGCCGCAAGGCGCCTG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGGGCTGCGCGGCGGGGAGCTGGCCGCTGCGCCGCCACGGCCGCAAGGCGCCTG 480

Qy    481 GGTGGCTCCCGCAGCCCGCGGC 500
          ||||||||||||||||||||||
Db    481 GGTGGCTCCCGCAGCCCGCGGC 500
```

RESULT 2
HSU28368
LOCUS   HSU28368   1322 bp   mRNA   linear   PRI 03-AUG-1995
DEFINITION   Human Id-related helix-loop-helix protein Id4 mRNA, complete cds.
ACCESSION   U28368
VERSION   U28368.1   GI:881545
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1322)
AUTHORS   Kiesling,T.L. and Christy,B.A.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1322)
AUTHORS   Kiesling,T.L.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUN-1995) Traci L. Kiesling, Institute of Biotech,
             UTHSCSA, 15355 Lambda Dr., San Antonio, TX 78245, USA
FEATURES   Location/Qualifiers
     source   1..1322
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /tissue_type="abdominal adipose tissue"
                     /dev_stage="15-year-old"
                     /note="Caucasian"

     CDS   376..861
                     /note="DNA binding protein inhibitor; Id-related

     misc_feature   565..687
                     /note="encodes helix-loop-helix region"

                     helix-loop-helix protein"
                     /codon_start=1
                     /product="Id4"
                     /protein_id="AAA73923.1"
                     /db_xref="GI:881546"
                     /translation="MKAVSPVRPSGRKAPSGGCGGGELALRCLAEHGHSLGGSAAAAA
                     AAAARCRAAEAAADEPALCLQCDMNDCYSRLRRLVPTIPPNKKVSKVEILQHVIDYIL
                     DLQLALETHPALLRQPPPPAPPHHPAGTCPAAPPRTPLITALNTDPAGAVNKQGDSIILC
                     R"

```
BASE COUNT      267 a      376 c      417 g      262 t
ORIGIN

Query Match                100.0%;  Score 500;  DB 9;  Length 1322;
Best Local Similarity      100.0%;  Pred. No. 9.3e-72;
Matches    500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGAATTCCGGGCGCCGGTTGTGAGTAGTACCGGGAGTGGGGTGATCCCGGGGCTAGGGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAATTCCGGGCGCCGGTTGTGAGTAGTACCGGGAGTGGGGTGATCCCGGGGCTAGGGGAGC 60

Qy     61 GCGGGCGCCCGATCGGGCGGCTTAGTGCGGAGCTCCGAAGGAGGACACCCGGGGTGGG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCGGGCGCCCGATCGGGCGGCTTAGTGCGGAGCTCCGAAGGAGGACACCCGGGGTGGG 120

Qy    121 CTACTTTTTCTTCCGGTGCTTTTGCTTTTTTGGGTCGGGCTTGAGTGTCGGTCGCCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTACTTTTTCTTCCGGTGCTTTTGCTTTTTTGGGTCGGGCTTGAGTGTCGGTCGCCC 180

Qy    181 ACTGAGACAAAAGATTCCCTCGTAAAACCCAGAGCGACACCCTCCCGTCAATTGTTGGGCTCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACTGAGACAAAAGATTCCCTCGTAAAACCCAGAGCGACACCCTCCCGTCAATTGTTGGGCTCGG 240

Qy    241 GAGTGTCGCGGTGCCCCGAGCGGGCGGGAGGCAAAGGAGGAGCCGGCCCGCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGTGTCGCGGTGCCCCGAGCGGGCGGGAGGCAAAGGAGGAGCCGGCCCGCG 300

Qy    301 GACGGGGGCCCGGAGCTTGGCCTCGCCTCGCGGCGGCGGGTTGCGCTCGCGGTAGAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACGGGGGCCCGGAGCTTGGCCTCGCCTCGCGGCGGCGGGTTGCGCTCGCGGTAGAG 360

Qy    361 CGCAGGGGGCGCGCGACTGAAGCGGGTGAGCCCCTCCGGCCCAAGGCGCCG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGCAGGGGGCGCGCGACTGAAGCGGGTGAGCCCCTCCGGCCCAAGGCGCCG 420

Qy    421 TCGGGCTGCGCGGCGGGGAGCTGGCCGCTGCGCCGCCACGGCCGCAAGGCGCCTG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCGGGCTGCGCGGCGGGGAGCTGGCCGCTGCGCCGCCACGGCCGCAAGGCGCCTG 480

Qy    481 GGTGGCTCCCGCAGCCCGCGGC 500
          ||||||||||||||||||||||
Db    481 GGTGGCTCCCGCAGCCCGCGGC 500
```

RESULT 3
BC014941
LOCUS   BC014941   2389 bp   mRNA   linear   PRI 29-OCT-2001
DEFINITION   Homo sapiens, inhibitor of DNA binding 4, dominant negative
             helix-loop-helix protein, clone MGC:20126 IMAGE:4552357, mRNA,
             complete cds.
ACCESSION   BC014941
VERSION   BC014941.1   GI:15928943
KEYWORDS   MGC.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2389)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer

```
REMARK        Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Genome Sequence Centre,
              BC Cancer Agency, Vancouver, BC, Canada
              info@bcgsc.bc.ca
              Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
              Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
              Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
              Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
              Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
              Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
              Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
              George Yang, Scott Zuyderduyn, Marco Marra.

              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              Series: IRAL Plate: 29 Row: b Column: 2
              This clone was selected for full length sequencing because it
              passed the following selection criteria: matched mRNA gi: 4504572.
FEATURES      Location/Qualifiers
     source        1..2389
                   /organism="Homo sapiens"
                   /db_xref="LocusID:3400"
                   /db_xref="taxon:9606"
                   /clone="MGC:20126 IMAGE:4552357"
                   /tissue_type="Uterus, leiomyosarcoma"
                   /clone_lib="NIH_MGC_46"
                   /lab_host="DH10B-R"
                   /note="Vector: pOTB7"
     CDS           370..855
                   /codon_start=1
                   /product="inhibitor of DNA binding 4, dominant negative
                   helix-loop-helix protein"
                   /protein_id="AAH14941.1"
                   /db_xref="GI:15928944"
                   /translation="MKAVSPVRPSGCRKAPSGCGGGELALRCLAEHGHSLGGSAAAAA
                   AAAARCKRAAEAAAADEPALCLQCDMNDCYSRLRLVPTIPNKKVSKVEILQHVIDYIL
                   DLQLALETHPALLRQPPPAPHHPAGTCPAAPPRTPLTALNTDPAGAVNKQGDSLLC
                   R"
BASE COUNT      636 a    510 c    592 g    651 t
ORIGIN

Query Match   95.5%;  Score 477.4;  DB 9;  Length 2389;
Best Local Similarity  99.4%;  Pred. No. 3.7e-68;
Matches 490;  Conservative  0;  Mismatches  1;  Indels  2;  Gaps  1;

Qy    10  GGCCGCGGTTGTGAGTAGTACCGGGAGTGGGGTGATCCCGGCTAGGGAGCGCGGCG--C  67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     2  GGCCGCGGTTGTGAGTAGTACCGGGAGTGGGGTGATCCCGGCTAGGGAGCGCGGCC   61

Qy    68  CCGATCGGGGGCTWAGTCGGGAGCTCCGAAGGAGGAGGAGGAGGAGGAGGAGGCTACTTT  127
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    62  GCGATCGGGGGCTWAGTCGGGAGCTCCGAAGGAGGAGGAGGAGGAGGAGGCTACTTT   121

Qy   128  TCTTCCGGTGCTTTGCTTTTTTTTTTTTTCCTTTGGCCTGTCTCGGGGCTCCCACTGAGC  187
          |||||||||||||| ||||||||||  |||||||||| |||||| |||||||||||||
Db   122  TCTTCCGGTGCTTTTGCTTTTTTTTTTCCTTTGGCGTGTCGCCCACTGAGC   181

Qy   188  AAAGATTCCTCGTAAAACCCAGAGAGGGACACCTCCCGTCAATTGTTGGGCTCGGGAGTGTC  247
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182  AAAGATTCCTCGTAAAACCCAGAGAGGGACAACCTCCCGTCAATTGTTGGGCTCGGGAGTGTC   241

Qy   248  GCGGTGCCCCGAGCGCCGCGCGCGGCGAAAAGGAGGGAGGGAGGCGGGG  307
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   242  GCGGTGCCCCAGCGCCGCGGCGAGGAGGCAAAGGGAGGCGGGG   301

Qy   308  CCCGGAGCCTTGCCTGCCTGCTCCGCCGCCGCTCCGGGTTCGCTTCGCTCGCGTCCTAGAGGCGCAGG  367
```

```
RESULT 4
HS625H18/c
LOCUS       HS625H18      185371 bp    DNA    linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from clone 625H18 on chromosome 6p22.2-23.
            Contains the gene for ID4 Helix-loop-helix DNA binding protein and
            part of an alternatively spliced novel gene. Contains ESTs, STSs,
            GSSs and putative CpG islands, complete sequence.
ACCESSION   AL022726
VERSION     AL022726.1  GI:3676217
KEYWORDS    HTG; CpG island; Helix-loop-helix; ID4.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185371)
  AUTHORS   Mashreghi-Mohammadi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 30, 1998 this sequence version replaced gi:3646106.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 625H18. The true right
            end of clone 498I24 (AL031057) is at 51918 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            625H18 is from the library RPCI14 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES      Location/Qualifiers
     source        1..185371
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="6"
                   /map="22.2-23"
                   /clone="RP4-625H18"
                   /clone_lib="RPCI-4"
     variation     259..261
                   /note="clone 498I24; ggc in this entry; substitution"
     variation     711
                   /note="clone 498I24; gtc"
                   /replace="gtc"
     repeat_region 711..758
                   /note="L2 repeat: matches 2660..2709 of consensus"
     variation     2117..2119
                   /note="clone 498I24; ttg in this entry; substitution"

Db   302  CCCGGAGCGTTGCCTCGCTCCGCTCCGCTCCGCCCCTCGCCTCCCCCACGCGGGTTCGCGTGCGGCCAGGG  361
          |||||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Qy   368  CCCGGCGGATGAAGGCGGTGAGCCCGGTCCTCGCCTCGCCCCCTCGCCTCCCCCACGCGGGTTGCGTGGGGCT  427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   362  CCCGCGCGGATGAAGGCGGTGAGCCCGGTCCTCGCCTCGCCCCCTCGCCTCCCCCACGCGGGTTGCGTGGGGCT  421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   428  GCGGGGCGGGGGAGCTGGCGCTGCGCTGCCTGGCCGAGCACGGCCACAGCCTGGGTGGCAGCGCGGGGGCT  487
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422  GCGGGGCGGGGGAGCTGGCGCTGCGCTGCCTGGCCGAGCACGGCCACAGCCTGGGTGGCAGCGCGGGGGCT  481

Qy   488  CCCGCAGCGCCGCGGC  500
          ||||||||||||||||
Db   482  CCCGCAGCGCCGCGGC  494
```

```
variation
        2889. .2891
        /replace="tcg"
        /note="clone 498I24; cat in this entry; substitution"
repeat_region
        3081. .3178
        /note="Trigger3(Golem) repeat: matches 1. .98 of consensus"
repeat_region
        3177. .3243
        /note="MIR1A2 repeat: matches 284. .351 of consensus"
variation
        3673. .3675
        /replace="cga"
        /note="clone 498I24; caa in this entry; substitution"
variation
        3774. .3776
        /replace="cga"
        /note="clone 498I24; ata in this entry; substitution"
repeat_region
        4028. .4177
        /note="L2 repeat: matches 2559. .2733 of consensus"
variation
        4045. .4047
        /replace="cgg"
        /note="clone 498I24; cag in this entry; substitution"
variation
        4524. .4526
        /replace="gca"
        /note="clone 498I24; ttt in this entry; substitution"
variation
        4651. .4653
        /replace="tct"
        /note="clone 498I24; cct in this entry; substitution"
variation
        4865. .4867
        /replace="cgt"
        /note="clone 498I24; gta in this entry; substitution"
variation
        5060. .5062
        /replace="gca"
        /note="clone 498I24; cgt in this entry; substitution"
variation
        5111. .5113
        /replace="agt"
        /note="clone 498I24; att in this entry; substitution"
repeat_region
        5210. .5615
        /note="L1MD2 repeat: matches 5748. .6164 of consensus"
variation
        5349. .5351
        /replace="cgg"
        /note="clone 498I24; cag in this entry; substitution"
variation
        5444. .5447
        /replace="atcg"
        /note="clone 498I24; actg in this entry; substitution"
variation
        5588. .5590
        /replace="tct"
        /note="clone 498I24; ttt in this entry; substitution"
variation
        5670. .5672
        /replace="cgt"
        /note="clone 498I24; cat in this entry; substitution"
variation
        5701. .5703
        /replace="gta"
        /note="clone 498I24; gca in this entry; substitution"
repeat_region
        5818. .6130
        /note="AluSq repeat: matches 1. .313 of consensus"
variation
        5916. .5918
        /replace="atg"
        /note="clone 498I24; acg in this entry; substitution"
misc_feature
        complement(6131. .6656)
        /note="match: GSSS AQ224537 B56744"
variation
        6211. .6213
        /replace="gcg"
        /note="clone 498I24; gtg in this entry; substitution"
variation
        6445. .6447
        /replace="cgt"
        /note="clone 498I24; cat in this entry; substitution"
variation
        6774. .6776
        /replace="gct"
        /note="clone 498I24; gtt in this entry; substitution"
variation
        6836. .6838
        /replace="tac"
        /note="clone 498I24; tgc in this entry; substitution"
variation
        6924. .6926
        /replace="aga"
        /note="clone 498I24; aaa in this entry; substitution"

variation
        6940. .6942
        /replace="tcc"
        /note="clone 498I24; ttc in this entry; substitution"
variation
        7288. .7291
        /replace="tcgg"
        /note="clone 498I24; tgtg in this entry; substitution"
variation
        7478. .7480
        /replace="agt"
        /note="clone 498I24; act in this entry; substitution"
variation
        7823. .7825
        /replace="gtc"
        /note="clone 498I24; gcc in this entry; substitution"
variation
        8324. .8326
        /replace="cat"
        /note="clone 498I24; cgt in this entry; substitution"
variation
        8478. .8480
        /replace="gcc"
        /note="clone 498I24; gtc in this entry; substitution"
variation
        8516. .8518
        /replace="gtc"
        /note="clone 498I24; gcc in this entry; substitution"
repeat_region
        8834. .8863
        /note="L2 repeat: matches 2714. .2743 of consensus"
variation
        9029. .9030
        /replace="cagaga"
        /note="clone 498I24; ca in this entry; deletion"
repeat_region
        9030. .9077
        /note="24 copies 2 mer ag 96% conserved"
variation
        9127. .9129
        /replace="ctt"
        /note="clone 498I24; cct in this entry; substitution"
variation
        9160. .9162
        /replace="ttg"
        /note="clone 498I24; tcg in this entry; substitution"
repeat_region
        9305. .9608
        /note="MIR repeat: matches 34. .258 of consensus"
variation
        9517. .9519
        /replace="cag"
        /note="clone 498I24; cgg in this entry; substitution"
variation
        9587. .9594
        /replace="aa"
        /note="clone 498I24; aaaaaaaa in this entry; insertion"
repeat_region
        9629. .9631
        /note="AluY repeat: matches 1. .304 of consensus"
repeat_region
        9822. .10484
        /note="MER6 repeat: matches 1. .697 of consensus"
variation
        10038. .10040
        /replace="acg"
        /note="clone 498I24; atg in this entry; substitution"
variation
        10056. .10058
        /replace="cgt"
        /note="clone 498I24; cat in this entry; substitution"
variation
        10171. .10173
        /replace="cat"
        /note="clone 498I24; cgt in this entry; substitution"
variation
        10219. .10220
        /replace="gcc"
        /note="clone 498I24; gc in this entry; deletion"
repeat_region
        10485. .10764
        /note="AluY repeat: matches 1. .285 of consensus"
variation
        10530. .10532
        /replace="ggg"
        /note="clone 498I24; ggg in this entry; substitution"
variation
        10606. .10608
        /replace="gcg"
        /note="clone 498I24; aca in this entry; substitution"
repeat_region
        10765. .10939
        /note="MER6 repeat: matches 697. .865 of consensus"
variation
        10878. .10880
```

RESULT 5

Query Match   95.5%;   Score 477.4;   DB 9;   Length 185371;
Best Local Similarity 99.4%;   Pred. No. 1.5e-68;
Matches 490;   Conservative 0;   Mismatches 1;   Indels 2;   Gaps 1;

```
Qy       10 GGCGCGCGGTTGTGAGTAGTACCGGGAGTGGGGTGGGGTGGATCCCGGGGCTAGGGAGGAGCGGCGCG--C  67
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db    72995 GGCGCGCGGTTGTGAGTAGTACCGGGAGTGGGGTGGGGTGGATCCCGGGGCTAGGGAGGAGCGGCGCGGCC  72936

Qy       68 CCGATCCGGGCTTAGTCGGAGCTCCGAAGGGAGTGGGGGAGCGGGGGGAGCGGCGCTGACTACTT  127
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72935 CCGATCCGGGCTTAGTCGGAGCTCCGAAGGGAGTGGGGGAGCGGGGGGAGCGGCGCTGACTT  72876

Qy      128 TCTTCCGGTGGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTCGGCCTGAGTGTCGGCCCCACTGAGC  187
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72875 TCTTCCGGTGGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTCGGCCTGAGTGTCGGCCCCACTGAGC  72816

Qy      188 AAAGATTCCCTCGTAAAACCCAGAAGGACGACCCTCCCGTCAATTGTTGGGGCTCGGGGAGTG  247
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72815 AAAGATTCCCTCGTAAAACCCAGAAGGACGACCCTCCCGTCAATTGTTGGGGCTCGGGGAGTG  72756

Qy      248 GCGGTGCCCCGACGCGCGCCGGGGAGGGGAGCGGAGCCCGGGCGGCCGGCGCCGGGGACGGGG  307
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72755 GCGGTGCCCCGACGCGCGCCGGGGAGGGGAGCGGAGCCCGGGCGGCCGGCGCCGGGGACGGGG  72696

Qy      308 CCCGGAGCTTGCCTGCCTTCCCTCGCTCCGCGCGGGTTTCGCTCGGCCGTAGAGCGCCAGGG  367
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    72695 CCCGGAGCTTGCCTGCCTTCCCTCGCTCCGCGCGGGTTTCGCTCGGCCGTAGAGCGCCAGGG  72636

Qy      368 CGGCGCCGATGAAGGCGGGTGAGCCCGGCGCAAGGCGCCGCTCGGGCT  427
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db    72635 CGGCGCCGATGAAGGCGGGTGAGCCCGGCGCAAGGCGCCGCTCGGGCT  72576

Qy      428 GCGCGCGCGGGACTGGCGGCCTGCCTCGCCTGCCGTCTGGGTGGCT  487
            ||||||||||||||||||||||||||||||||||||||||| |||||
Db    72575 GCGCGCGCGGGACTGGCGGCCTGCCTCGCCTGCCGTGGGGTGGCT  72516

Qy      488 CCGCAGCCGCGGGC  500
            ||||||||||||||
Db    72515 CCGCAGCCGCGGGC  72503
```

LOCUS   HSU16153
HSU16153   1017 bp   mRNA   linear   PRI 05-DEC-1995
DEFINITION Human Id-4H protein mRNA, complete cds.
ACCESSION   U16153
VERSION   U16153.1 GI:625095
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1017)
  AUTHORS   Pagliuca,A., Bartoli,P.C., Saccone,S., Della Valle,G. and Lania,L.
  TITLE   Molecular cloning of ID4, a novel dominant negative
            helix-loop-helix human gene on chromosome 6p21.3-p22
  JOURNAL   Genomics 27 (1), 200-203 (1995)
  MEDLINE   95394461
REFERENCE   2 (bases 1 to 1017)
  AUTHORS   Lania,L.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-OCT-1994) Luigi Lania, Genetica, Biologia Generale e
            Molecolare, University of Naples, Via Mezzocannone 8, Naples, 80134
            Italy
FEATURES   Location/Qualifiers
     source   1..1017
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3-p22"
     CDS   306..782
                     /codon_start=1
                     /product="Id-4H protein"
                     /protein_id="AAA82882.1"
                     /db_xref="GI:625096"
                     /translation="MKAVSPVRPRPLRPSGCCGGGEBLALRCLAEHGHSLGSQAAAAA
                     AARCKAAEAAAAADRPALCLQCDMNDCYSRLRWLPTTPPNKKVSKVEILQHVIDYLQ
                     LALETHPALLRQPPPAPPHHPAGTCPAAEPRTPLTLALNTDPAGAVNKQGDSILCR"

BASE COUNT   192 a   324 c   335 g   166 t
ORIGIN

RESULT 6

Query Match   63.2%;   Score 316.2;   DB 9;   Length 1017;
Best Local Similarity 92.2%;   Pred. No. 4.6e-42;
Matches 401;   Conservative 0;   Mismatches 23;   Indels 11;   Gaps 6;

```
Qy       66 GCCCGATCCGGGCTTAGTCGCAGCTCCGAAGGGAGTGGAGTGGGAACACCCGGGGTGGGGTGGGGCTACT  125
            ||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db        4 GCCGCATCCGGGCTTAGTCGGAGCTCCGAA-GGAGTGACTAGGACACCCGGGGTGGGGCTACT  62

Qy      126 TTTCTTCCGGTGGCTTTTGCTTTTTTTTTTTTTCCTTTGGGCTCGGCCTGAGTGTCGCCCACTGA  185
            ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       63 TTTCTTCCGGTGGCTTTTTGCT-TTTTTTTTTCCTTTGGGCTCGGCCTGAGTGTCGCCCACTGA  121

Qy      186 GCAAAGATTCCCTCGTAAAACCCAGAAGGACGACCCTCCCGTCAATTGTTGGGGCTCGGGGAGTG  245
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      122 GCAAAGATTCCCTCGTAAAACCCAGAGGACGACCCTCCCGTCAATTGTTGGGGCTCGGGGAGTG  181

Qy      246 TCGCGCGGTGCCCCGAGGCGCGCGGGGAGGCAAAGGGAGGGGAGCGGCGCCGGGGACGG  305
            ||||||||||||||||||||||||||||          ||||||||||||||||||||||
Db      182 TCGCGCGGTGCCCCGAGGCGCGCGGGGC---CAGGCAAAGGGAGGGGA--CGGGGACGG  237

Qy      306 GGCCCCGGAGCTTGCCTGCCTTCCCTCGCTCCGCGCGGGTTCGCTCCGTAGAGCGCCAG  365
            ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      238 GGCCCCGGAGCTTGCCTGCCTTCCCTCGCTCCGCGCGGGGTTCGCTCCGCGCAGGGCG--A  295

Qy      366 GGCGCGCCGCGGGACGGTGAGCGCGGGTGAGCCCGGCGCCCCCTTCGGGCCGCAAGGCGCCGTCGGG  425
            ||          ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db      296 GGGCGCGCGCGGATGAAAGCGGGTGGTGAGCCCGGCGCCCCCTTCGGGGCTCGGGGCTCGGG  352

Qy      426 CTGCGCGCGCGGGACGTGGCGGCTGCGCCGTCGCCTGCCTGGCCCGGAGCCGCACCAGCCTGGGGTGG  485
            |||| ||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db      353 CTGCGCGCGCGGGAGACGTGGCGGCTGAGCCGGGTGGCCCTGCCTAA--GGCCGTCGGG  412

Qy      486 CTCCGCAGCCGCGGGC  500
            ||| |  |||||
Db      413 CTCGCAGCCGCGGGC  427
```

LOCUS   AF030295
AF030295   1473 bp   DNA   linear   PRI 02-APR-1998
DEFINITION Homo sapiens ID4 gene, promoter region.
ACCESSION   AF030295
VERSION   AF030295.1 GI:3004879
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1473)
  AUTHORS   Pagliuca,A., Cannada-Bartoli,P. and Lania,L.
  TITLE   A role for Sp and helix-loop-helix transcription factors in the
            regulation of the human Id4 gene promoter activity
  JOURNAL   J. Biol. Chem. 273 (13), 7668-7674 (1998)
  MEDLINE   98184879
REFERENCE   2 (bases 1 to 1473)
  AUTHORS   Lania,L.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-OCT-1997) Dipartimento di Genetica, Biologia Generale
            e Molecolare, Universite degli Studi di Napoli 'Federico II', via
            Mezzocannone 8, Napoli, NA 80134, Italy
FEATURES   Location/Qualifiers
     source   1..1473
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"

```
BASE COUNT       287 a      431 c      408 g      347 t
ORIGIN

FEATURES             Location/Qualifiers
     source          1..1473
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"

SOURCE          synthetic construct.
  ORGANISM      synthetic construct
                artificial sequence.
REFERENCE       1  (sites)
  AUTHORS       Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE         Diagnosis of diseases associated with dna transcription
  JOURNAL       Patent: WO 0192565-A 335 06-DEC-2001;
                Epigenomics AG (DE)

KEYWORDS    .
VERSION     AX323847.1    GI:18094596
ACCESSION   AX323847
DEFINITION  Sequence 335 from Patent WO0192565.
LOCUS       AX323847              1473 bp    DNA        linear    PAT 07-JAN-2002

RESULT 7

Qy        1466 GCGCG 1470
             |||||
Db        1466 GCGCG 1470

Qy         369 GCGCG 373
             |||||
Db        1406 CCGGAGCGTTGCCTTGCCTCCCTCGCTCCCCCAGCCGGGGTTCGCCGGTTGGGGACGGGGC 1465

Qy         309 CCGGAGCGTTGCCTTGCCTCCCTCGCTCCCCCAGCCGGGGTTCGCCGGTTGGGGACGGGGC 368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1406 TCGGAGTTTGTTTGTTTTGTTTTCGTTCGTTTTACCGGGTTCGTTCGCGTTAGGGCGG 1465

Qy         309 CCGGAGCGTTGCCTTGCCTCCCTCGCTCCCCCAGCCGGGGTTCGCCGGTTGGGGACGGGGC 368
             ||||  ||||||||  |||  |||||  ||||  |||||  ||||||||  |||||
Db        1350 CGGTTGTTTCGAGCGCGTCGGG---TTAGGTAAAGGGAGCGGA-TCGGTCGCGACGGGGT 1405

Qy         249 CGGTGCCCGAGCGCGCGGGCGCGGGGAGGGAGGGAGGGAGGGAGGGAGCCTCCGTCAATTGTTGGGCTTCGGAGTGTCG 308
             |||||||  ||||||||||  ||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        1290 AAGAATTCCCTCGTAAAAACCCAGAGAGCGACCTCCCGTCAATTGTTGGGCTTCGGAGTGTCG 1349

Qy         189 AAGAATTCCCTCGTAAAAACCCAGAGAGCGACCTCCCGTCAATTGTTGGGCTTCGGAGTGTCG 248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1231 CTTCCGGTGCTTTCGTTTTTGTTTTTTTTCCTTTGGGCTTCGGAGTGTTCGCCCACTGAGCA 1289

Qy         129 CTTCCGGTGCTTTCGTTTTTGTTTTTTTTCCTTTGGGCTTCGGAGTGTTCGCCCACTGAGCA 188
             ||||  |||  |||||  ||||||  |||  ||||||||||  ||||||||  |||  ||||
Db        1172 GCATCGGGCTTAGTCGGAGCGTCCGAA-GGAGTGACTAGGACACCCGGGTTATTTTT 1230

Qy          70 G-ATCGGGCTTAGTCGGAGCGTCCGAAGGGAGTGACTAGGACACCCGGGTTATTTTT 128
             ||  ||||||||||||||||  ||||||  ||||||  ||||  |||||||||||
Db        1112 GGCCGCGGGTTGTGAGTAGTACCGGGAGTGGGGGTGATCCGGGCTCGGGCCTACTTT 1171

Qy          10 GGGCGCGGGTTGTGAGTAGTACCGGGAGTGGGGGTGATCCGGGCTCGGGCCTACTTT 69
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         287 GGGCGCGGGTTGTGAGTAGTACCGGGAGTGGGGGTGATCCGGGCTCGGGCCTACTTT 346

Matches 350; Conservative    0; Mismatches    8; Indels    7; Gaps    5;
Best Local Similarity   95.9%; Conservative   0;
Query Match       57.6%;        Score 288.2;  DB 9;   Length 1473;
```

```
misc_feature   1..1470
               /note="this region contains promoter and 5' end of mRNA"
TATA_signal    1066..1071
CDS            1471..>1473
               /codon_start=1
               /product="ID4"
               /protein_id="AAC39674.1"
               /db_xref="GI:3004880"
               /translation="M"
               /map="6p21.3-22"
```

```
REFERENCE    1  (bases 1 to 176835)
  AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
             Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
             Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
             Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.J.,
             Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
             Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
             Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
             Hamilton,K., Harris,K., Hart,M., Havlak,P., Hawes,A.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
             Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
             Martinez,E., Massey,E., Mawhiney,E., McLeod,M.,
             Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
             Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
             Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
             Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
             Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
```

```
RESULT 8

Qy        1466 GCGCG 1470
             |||||
Db        1466 GCGCG 1470

Qy         369 GCGCG 373
             |||||
Db        1406 TCGGAGTTTGTTTGTTTTGTTTTCGTTCGTTTTACCGGGTTCGTTCGCGTTAGGGCGG 1465

Qy         309 CCGGAGCGTTGCCTTGCCTCCCTCGCTCCCCCAGCCGGGGTTCGCCGGTTGGGGACGGGGC 368
             ||||  ||  ||||||  ||||  ||||  ||||||  ||||||||  |||
Db        1350 CGGTTGTTTCGAGCGCGTCGGG---TTAGGTAAAGGGAGCGGA-TCGGTCGCGACGGGGT 1405

Qy         249 CGGTGCCCGAGCGCGCGGGCGCGGGGAGGGAGGGAGGGAGGGAGGGAGGCCTCCGTCAATTGTTGGGCTTCGGAGTGTCG 308
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1290 AAGAATTCCCTCGTAAAAACCCAGAGAGCGACCTCCCGTCAATTGTTGGGCTTCGGAGTGTCG 1349

Qy         189 AAGAATTCCCTCGTAAAAACCCAGAGAGCGACCTCCCGTCAATTGTTGGGCTTCGGAGTGTCG 248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1231 CTTCCGGTGCTTTCGTTTTTGTTTTTTTTCCTTTGGGCTTCGGAGTGTTCGCCCACTGAGCA 1289

Qy         129 CTTCCGGTGCTTTCGTTTTTGTTTTTTTTCCTTTGGGCTTCGGAGTGTTCGCCCACTGAGCA 188
             |||  |||||||||||  |||  ||||||  |||||||  |||||||  ||||||
Db        1172 GTATCGGGCTTAGTCGGAGTTTCGAA-GGAGTGATTAGGATATTCGGGTGTGGGTTTATTTTT 1230

Qy          70 G-ATCGGGCTTAGTCGGAGCGTCCGAAGGGAGTGACTAGGACACCCGGGTTATTTTT 128
             |  ||||||||||  ||||||  ||  ||||||  ||  |||||||||||
Db        1112 GGCCGCGGGTTGTGAGTAGTACCGGGAGTGGGGGTGATCCGGGCTCGGGCCTACTTT 1171

Qy          10 GGGCGCGGGTTGTGAGTAGTACCGGGAGTGGGGGTGATCCGGGCTCGGGCCTACTTT 69
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         287 GGGCGCGGGTTGTGAGTAGTACCGGGAGTGGGGGTGATCCGGGCTCGGGCCTACTTT 346

Matches 288; Conservative    0; Mismatches   70; Indels    7; Gaps    5;
Best Local Similarity   78.9%; Pred. No. 1.5e-21;
Query Match       37.8%;        Score 189;  DB 6;   Length 1473;
```

```
SOURCE      Rattus norvegicus
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
KEYWORDS    HTG; HTGS_PHASE1.
VERSION     AC099470.2    GI:17974445
ACCESSION   AC099470
DEFINITION  Rattus norvegicus clone CH230-221D3, ***  SEQUENCING IN PROGRESS
            ***, 84 unordered pieces.
LOCUS       AC099470              176835 bp    DNA        linear    HTG 20-DEC-2001
```

```
BASE COUNT       287 a      111 c      408 g      667 t
ORIGIN

Matches 288; Conservative    0; Mismatches   70; Indels    7; Gaps    5;
Best Local Similarity   78.9%; Pred. No. 1.5e-21;
Query Match       37.8%;        Score 189;  DB 6;   Length 1473;
Query Match
```

```
AUTHORS   Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
          Ruiz,S., Savery,G., Scherer,S., Scott,D., Shen,H., Shooshtari,N.,
          Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
          Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
          Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
          Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
          Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
          Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
          Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 176835)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Dec 20, 2001 this sequence version replaced gi:16930877.
          ------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          ------- Project Information
          Center project name: GIZO
          Center clone name: CH230-221D3
          ------- Summary Statistics
          Assembly program: Phrap; version 0.990329First call to
          findPhrapList
          Consensus quality: 131300 bases at least Q40
          Consensus quality: 138970 bases at least Q30
          Consensus quality: 145929 bases at least Q20
          Estimated insert size: 130110; sum-of-contigs estimation
          Quality coverage: 0x in Q20 bases; agarose-fp estimation
          Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
          --------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE:This is a 'working draft' sequence. It currently
          * consists of 84 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *            1
          *         6956: contig of 6956 bp in length
          *         6957
          *         7056: gap of unknown length
          *         7057
          *        12271: contig of 5215 bp in length
          *        12272
          *        12371: gap of unknown length
          *        12372
          *        16946: contig of 4575 bp in length
          *        16947
          *        17046: gap of unknown length
          *        17047
          *        21276: contig of 4230 bp in length
          *        21277
          *        21376: gap of unknown length
          *        21377
          *        25033: contig of 3657 bp in length
          *        25034
          *        25133: gap of unknown length
          *        25134
          *        28831: contig of 3698 bp in length
          *        28832
          *        28931: gap of unknown length
          *        28932
          *        32908: contig of 3977 bp in length
          *        32909
          *        33008: gap of unknown length
          *        33009
          *        36651: contig of 3643 bp in length
          *        36652
          *        36751: gap of unknown length
          *        36752
          *        39620: contig of 2869 bp in length
          *        39621
          *        39720: gap of unknown length
          *        39721
          *        43480: contig of 3760 bp in length
          *        43481
          *        43580: gap of unknown length
          *        43581
          *        47623: contig of 4043 bp in length
          *        47624
          *        47723: gap of unknown length
          *        47724
          *        49735: contig of 2012 bp in length
          *        49736
          *        49835: gap of unknown length
          *        49836
          *        52089: contig of 2254 bp in length
          *        52090
          *        52189: gap of unknown length
          *        52190
          *        54665: contig of 2476 bp in length
          *        54666
          *        54765: gap of unknown length
          *        54766
          *        57406: contig of 2641 bp in length
          *        57407
          *        57506: gap of unknown length
          *        57507
          *        60153: contig of 2647 bp in length
          *        60154
          *        60253: gap of unknown length
          *        60254
          *        63252: contig of 2999 bp in length
          *        63253
          *        63352: gap of unknown length
          *        63353
          *        65370: contig of 2018 bp in length
          *        65371
          *        65470: gap of unknown length
          *        65471
          *        68011: contig of 2541 bp in length
          *        68012
          *        68111: gap of unknown length
          *        68112
          *        70803: contig of 2692 bp in length
          *        70804
          *        70903: gap of unknown length
          *        70904
          *        72986: contig of 2083 bp in length
          *        72987
          *        73086: gap of unknown length
          *        73087
          *        74859: contig of 1773 bp in length
          *        74860
          *        74959: gap of unknown length
          *        74960
          *        76202: contig of 1243 bp in length
          *        76203
          *        76302: gap of unknown length
          *        76303
          *        77797: contig of 1495 bp in length
          *        77798
          *        77897: gap of unknown length
          *        77898
          *        80646: contig of 2749 bp in length
          *        80647
          *        80746: gap of unknown length
          *        80747
          *        82795: contig of 2049 bp in length
          *        82796
          *        82895: gap of unknown length
          *        82896
          *        84216: contig of 1321 bp in length
          *        84217
          *        84316: gap of unknown length
          *        84317
          *        86357: contig of 2041 bp in length
          *        86358
          *        86457: gap of unknown length
          *        86458
          *        86515: contig of 2058 bp in length
          *        86516
          *        86615: gap of unknown length
          *        86616
          *        88515: contig of 1301 bp in length
          *        88516
          *        88615: gap of unknown length
          *        88616
          *        88916: contig of 1966 bp in length
          *        88917
          *        89016: gap of unknown length
          *        89017
          *        89916: contig of 1193 bp in length
          *        89917
          *        90016: gap of unknown length
          *        90017
          *        92523: contig of 2507 bp in length
          *        92524
          *        92623: gap of unknown length
          *        92624
          *        94248: contig of 1625 bp in length
          *        94249
          *        94348: gap of unknown length
          *        94349
          *        95832: contig of 1484 bp in length
          *        95833
          *        95932: gap of unknown length
          *        95933
          *        97898: contig of 1966 bp in length
          *        97899
          *        97998: gap of unknown length
          *        97999
          *        99191: contig of 1193 bp in length
          *        99192
          *        99291: gap of unknown length
          *        99292
          *       101227: contig of 1936 bp in length
          *       101228
          *       101327: gap of unknown length
          *       101328
          *       103172: contig of 1845 bp in length
          *       103173
          *       103272: gap of unknown length
          *       103273
          *       104839: contig of 1567 bp in length
          *       104840
          *       104939: gap of unknown length
          *       104940
          *       106467: contig of 1528 bp in length
          *       106468
          *       106567: gap of unknown length
          *       106568
          *       108405: contig of 1838 bp in length
          *       108406
          *       108505: gap of unknown length
          *       108506
          *       110280: contig of 1775 bp in length
          *       110281
          *       110380: gap of unknown length
          *       110381
          *       112198: contig of 1818 bp in length
          *       112199
          *       112298: gap of unknown length
          *       112299
          *       113542: contig of 1244 bp in length
          *       113543
          *       113642: gap of unknown length
          *       113643
          *       114787: contig of 1145 bp in length
          *       114788
          *       114887: gap of unknown length
          *       114888
          *       116687: contig of 1800 bp in length
          *       116688
          *       116787: gap of unknown length
          *       116788
          *       118703: contig of 1916 bp in length
          *       118704
          *       118803: gap of unknown length
          *       118804
          *       120736: contig of 1933 bp in length
          *       120737
          *       120836: gap of unknown length
          *       120837
          *       122508: contig of 1672 bp in length
          *       122509
          *       122608: gap of unknown length
          *       122609
          *       124119: contig of 1511 bp in length
          *       124120
          *       124219: gap of unknown length
          *       124220
          *       126465: contig of 2246 bp in length
```

```
REFERENCE    2 (bases 1 to 4701)
  AUTHORS    Sablitzky,F.
  JOURNAL    Submitted (08-OCT-1997) Sablitzky F., Medicine, The Windeyer
             Institute of Medical Sciences, University College London, 46
             Cleveland Street, London, W1P 6DB, UK
  TITLE      Direct Submission
REFERENCE    1 (bases 1 to 4701)
  AUTHORS    Sablitzky,F.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
SOURCE       house mouse.
KEYWORDS     Id4 gene.
VERSION      AJ001972.1  GI:2546999
ACCESSION    AJ001972
DEFINITION   Mus musculus Id4 dominant negative helix-loop-helix gene.
LOCUS        MMLH1972              4701 bp    DNA    linear   ROD 16-DEC-1998
MMLH1972

RESULT 9

Db  135409  GGC  135411
Qy     498  GGC  500
            |||
Db  135350  GGAGCTGGGCTACGTTGCTGGCGGGGAGCA-GGCCATAGACTAAGTGGCTCGGCAGCCCC  135408
            |||||||||||||||||||||||||||||  ||||||||||||| |||||||||||||
Qy     438  GGAGCTGGGCTGGCGCTGCCGCTGGCGGGGCCGAGCACGCGGCACGGGGTGGGCTCCGCAGCCGC  497
            ||||||||||  |||||||||||||||||||
Db  135290  GAAGGCGGTGAGCCGGTGGCCCCCTCTCGGCCCGCCGTGTCGGGGCTGCGGCGGGGGG  135349
            |||||||||  .
Qy     378  GAAGGCGGTGAGCCGCCGGTGCGCCGGCCAAGGCGGGCCCCCCCGCGGCGGGGGCGGGG  437
            ||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db  135239  CGCT--CTACCGCTTGTCCCCCGCGGGGT------CCTCGGCGCAGGAAGCGCGCGAT  135289
            |||||     ||||| |||       |  ||||||  ||||||||  ||||
Qy     318  GCCTGCTTCCCTCGCGTCGCCCCCAGCCGGTTCGCTCCGTAGAGCGCAGGCGCCGCGGAT  377
            |||||||||||||||||||||||||||||||  |||||||||||||||||
Db  135181  GA--GCGGCCGGCGCCGAGGGGGAGCAAAGCGTATCGGGCCGGGTACCGGAGCT  135238
            ||  |||||||||||||||||||||||  |||||||||||||||||||||
Qy     258  GAGCGCGTCCCTCGCTCGCCCCCAGCGGGGTTCGCTCGCCGCAGGGGCCCGGGAGCTT  317
            |||||||||||||||||||||||||||||||||||||||||||||
Db  135121  GCCCAAAACCCCGAGCAGCCCCTCCGGTTCGAGTCTTGGAACCGCGCCGGGGTGCTGC  135180
            |||||||||||||||||||||||||    |||||||||||||
Qy     202  -----AAAACCCAGAGACGGACGACCCTCGCAGTCAAATTGTTTGGGCGTGTCGCGGTGCCCC  257
            ||||||  |||| ||||||||||||||||||||||
Db  135061  TCATTTTTCTTGTGTCTTTTTGCTTCGGGCCGAGTCGCTCACTAAAAACTGAGAGATTCCCTCGT  135120
            |||||||   ||||||  |||||||||||||||||||||||||||||
Qy     142  TGCTTTTTTTTTTTTCCTTTGGGCTTCGGGGTCTGCCGCCCACGCAGCAAAGATTCCCTCGT  201
            |||||||||||||||  |||||||||||||||||
Db  135001  TCGAAAGAGAAACGGCTCAGGGACCACAGGGTGGGCTACTTTGTTGTTTTTTTTCTTTACCGGCTTT  135060
            |||||||||||||||||||||     |||||||||  |||||||||||||
Qy      89  TCCGAAGGGAGGAGTGACTAGGACACCCGGGTGGGCTAC-------TTTTCTTCCGGTGCTTT  141
            |||||||||||||||||| |||||||||||||||||||       ||||||
Db  134941  GAGGGGGGGGGCAGTCACCGGCTAAGGGAGAGTCGGGGACAGCGGATCCACCTTAGTCGGAGC  135000
            ||||||||||||||||| |||||||||||||||||||||  ||||||||
Qy      31  GGGGAGTGGGGTGGGATCCCGGGCTAGGGAGGAGCGCGCGCC--CGATCGGGGCTTAGTCGGAGC  88
```

Matches 342; Conservative 70.8%; Pred. No. 0; Mismatches 116; Indels 25; Gaps 7;
Best Local Similarity 35.5%; Score 177.4; DB 2; Length 176835;
Query Match

```
Db  135533  135632:    gap of unknown length
Db  135533  135532:    contig of 1779 bp in length
Db  133754  133753:    gap of unknown length
Db  133654  133653:    contig of 1331 bp in length
Db  132323  132322:    gap of unknown length
Db  131057  131056:    contig of 1166 bp in length
Db  130957  130956:    contig of 1291 bp in length
Db  129566  129565:    gap of 1775 bp in length
Db  127791  127790:    gap of unknown length
Db  127691  127690:    contig of 1125 bp in length
Db  126466  126565:    gap of unknown length
```

AUTHORS   vanCruechten,I., Cinato,E., Fox,M., King,E.R., Newton,J.S.,
           Riechmann,V. and Sablitzky,F.
TITLE   Nucleotide Sequence, Structure, Chromosomal Localisation,
           Transcription and Expression Pattern of the Murine Id4 Gene
JOURNAL   Unpublished
REFERENCE 3 (bases 1 to 4701)
  AUTHORS Riechmann,V. van Cruchten,I. and Sablitzky,F.
  TITLE   The expression pattern of Id4, a novel dominant negative
           helix-loop-helix protein, is distinct from Id1, Id2 and Id3
  JOURNAL Nucleic Acids Res. 22 (5), 749-755 (1994)
  MEDLINE 94188125
REFERENCE 4 (bases 1 to 4701)
  AUTHORS van Cruchten,I., Cinato,E., Fox,M., King,E.R., Newton,J.S.,
           Riechmann,V. and Sablitzky,F.
  TITLE   Structure, chromosomal localisation and expression of the murine
           dominant negative chromosomal helix-loop-helix Id4 gene
  JOURNAL Biochim. Biophys. Acta 1443 (1-2), 55-64 (1998)
  MEDLINE 99057583
COMMENT   1-1052 promoter,   1053-1493 exon I,   1494-1837 intron I,
           exon II, 1901-2504 intron II, 2505-4701 3 prime untranslated region
           (the 3 prime untranslated region of the Id4 gene continuous for about 1.8
           kb).

FEATURES   Location/Qualifiers
     source   1..4701
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
     exon   1053..1493
                     /gene="Id4"
                     /number=1
     CDS   join(1053..1493,1838..1882)
                     /codon_start=1
                     /protein_id="CAA05120.1"
                     /db_xref="SWISS-PROT:P41139"
                     /db_xref="GI:2547000"
                     /translation="MKAVSPVRPSGRKAPSGCGGGELALRCLAEHGHSLGGSAAAAAA
                     AAAARCKAAFAAADEPALCLQCDMNCYSRLRRLVPPIPPNKKVSKVEILQHVIDYIL
                     DLQLALETHPALLRQPPPAPLHPAGACPVAPPRTPLTALNIDPAGAVNKQGDSILC
                     R"
     gene   1053..1882
                     /gene="Id4"
     exon   1838..1882
                     /gene="Id4"
                     /number=2

BASE COUNT   1085 a 1190 c 1194 g 1232 t
ORIGIN

Matches 310; Conservative 67.0%; Pred. No. 0; Mismatches 137; Indels 16; Gaps 4;
Best Local Similarity 34.4%; Score 171.8; DB 10; Length 4701;
Query Match

```
Db  955  CGCCGGACCGCGGGGAGGGAGGGGACCAAAGTAAAATTGCGCGGCGGGGGGAGGACCCCGAGCGCTC  1014
         |||                ||||||     ||||   ||||  |||||
Qy  279  AAAGGAGCGGAGCGGCCGCCGCGACGGGGCCCGGGGAGCTTGCCTGCTCGGGGCTGCGGG  338
         |||||||||||||||||||  |||||||||||||||||||||||||||||
Db  895  CAAAACCCCGAGCAGCTCTCCGGTCGCGATTTCTCGGAGCTCGGAGCCGCGGGCCGGGGC  954
         |||||||||||||  |||||||||||||||||||||||  |||||
Qy  223  GTCAA---TTGTTGGGCTCGGGAGTGTCGCGGTGCCCCAGCCGGTCCTTC  278
         |||||   ||||||||||||||||||||||||||||||||||||
Db  835  CGATTGTCGCTCACAAAAAAGAAAAAACAAAAAACCTGAGATACCCTTTC  894
         |||||||||||||||||||||||||||||
Qy  169  TGAGTGTCGCGCCAC------TGAGCAAAGAATTCCCTCGTAAAACCCAGAGCGCAACCCTCCC  222
         |||||||||||||          |||||||||||||  ||||||||||||||||
Db  775  GGGTGGGCTACTTTTTTTTTATTTCCGGGGTCTTTTGATTTTTTCTTTGTCTTTTTGGG  834
         |||||||||||||           |||||||||||||||
Qy  114  GGGTGGGCTTACTTTTTC-----TTCCGGTGCTTTTTGCTTTTTTTTTTCCTTTGGGCTGGGC  168
         |||||||||||||        |||||||||  |||||  ||||||
Db  715  GGGGCAGTCACCGGCGGCGGATCCACCTTAGTCGAAGCTCAGGAGGACCA  774
         |||||||||||  ||||||||||||||||||||
Qy   54  GGGGAGCGCGCGGCGCCGATCGGGCGTTAGTCGAGCTCCGAGAGGGAGTGACTAGGACACCC  113
```

RESULT 10

```
AF077859

LOCUS       AF077859        4503 bp    DNA     linear   ROD 16-JAN-1999
DEFINITION  Mus musculus helix-loop-helix protein Id4 gene, complete cds.
ACCESSION   AF077859
VERSION     AF077859.1  GI:3420736
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4503)
  AUTHORS   Mantani,A., Hernandez,M.C., Kuo,W.L. and Israel,M.A.
  TITLE     The mouse Id2 and Id4 genes: structural organization and
            chromosomal localization
  JOURNAL   Gene 222 (2), 229-235 (1998)
  MEDLINE   9905133
REFERENCE   2  (bases 1 to 4503)
  AUTHORS   Mantani,A., Hernandez,M.C., Kuo,W.-L. and Israel,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-1998) Neurological Surgery, University of
            California, San Francisco, 515 Parnassus Ave., San Francisco, CA
            94143, USA
FEATURES             Location/Qualifiers
     source          1..4503
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
     CDS             join(1966..2406,2753..2797)
                     /codon_start=1
                     /product="helix-loop-helix protein Id4"
                     /protein_id="AAD05213.1"
                     /db_xref="GI:3420737"
                     /translation="MKAVSPVRPSGRKAPSGCGGGELALRCLAEHGHSLGGSAAAAAA
                     AAAARCKAAEAAADEPALCLQCDMNDCYSRLRRLVPTIIPNKKVSKVEILQHVIDYIL
                     DLQIALETHPALLRQPPPAPPLHPAGACVPAPPRTPLTALNTDPAGAVNKQGDSILC
                     R"
BASE COUNT    993 a  1183 c  1175 g  1152 t
ORIGIN
```

Best Match
Qy / Db alignment

```
Qy    339  CAGCGGGGT-TCGCTCGGGTAGAGCGCAGGGGGTGAGCGCCGGTGC 397
                    |||||||| |||||||||||||||||||||||||||||||||
Db   1015  TACCGGCTTGTCGCGGTTCTCTCCGCAGGAAGCGCGCGGTGAGCCCCGGTGC 1074

Qy    398  GCCCCTTCGGGGCGGCGGGGGGGGGGGGAGCTGGGCGTCGC 457
           |||||||||||||||||||||||||||||||||||||||||
Db   1075  GCCCCTCGGGCGCAAGGCGCGCGGGGGGGGGACTGGGCGTCGC 1134

Qy    458  TGGCCGGACCACGGCCCACCAGCCTGGGTTGGCTCCCGCAGCCCGGGC 500
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1135  TGGCCGGAGCACGGCCCACCAGCCTGGGTGGGCTCCCGCAGCCCCGC 1177
```

```
Qy     54  GGGGAGCGCGCGGCCCGATCGGCGCTTAGTCGGGAGGGAGTGACTAGGACACCC 113
                  |||||||||||||||||||||||||||||||||||||||
Db   1626  GGGGCAGTCACCGGGCGGATCCACCTTAGTCGAAGCTCAGAGGGAACGGCTCAGGGACCA 1685

Qy    114  GGGGTGGGGCTACTTTTC-----TTCCGCGTGCTTTTGCTTTTTTTTTTTTCCTTTGGGCTTCGGGC 168
           |||||||||||||       |||||||||||||||||||||||||||||||||||||||
Db   1686  GGGGTGGGGCTACTTTTTTTTTATTTCCGGGCTTTTGATTTTTCTTTGTTTTTTCGGGC 1745

Qy    169  TGAGTGTGCCCCACTGTTTGGGCTCGTAAAAGAATTCCCTCGTAAAAAACCCAGAGC-- 216
           ||||||||| ||||||||||||  ||||||||||||||||||||||||||||||
Db   1746  CGATGTCGCGCTCACAAAAAAGAAAAAAAAAAAACAAAAAACCTTGAGATACCCTTTC-GAC 1805

Qy    217  CCTCCGGTCAATTGTTGGGCTGGTGTGTCCCCCAGCGCCGCCCGGGGTGCTCGGCA 276
                   ||| |||||  |||||||||||||||||||||||||||||||||||||||
Db   1806  CAAAAACCCGAGCAGCTTCTCCGGTCGATTCTCGGGACTCGAGCGGCCGGGTGCTCGGCGA 1865
```

Query Match 34.0%; Score 170.2; DB 10; Length 4503;
Best Local Similarity 65.4%;
Matches 304; Conservative 0; Mismatches 143; Indels 18; Gaps 3;

RESULT 11

```
HSID4

LOCUS       HSID4       1147 bp    mRNA     linear   PRI 15-JAN-1999
DEFINITION  H.sapiens mRNA for ID4 protein.
ACCESSION   Y07958
VERSION     Y07958.1  GI:1550769
KEYWORDS    helix-loop-helix protein; Id4 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1147)
  AUTHORS   Rigolet,M., Rich,T., Gross-Morand,M.S., Molina-Gomes,D.,
            Viegas-Pequignot,E. and Junien,C.
  TITLE     cDNA cloning, tissue distribution and chromosomal localization of
            the human ID4 gene
  JOURNAL   DNA Res. 5 (5), 309-313 (1998)
  MEDLINE   99087490
REFERENCE   2  (bases 1 to 1147)
  AUTHORS   Rigolet,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1996) M. Rigolet, INSERM U383, Clinique.M.Lamy,
            Hospital Necker, 75015 PARIS, FRANCE
COMMENT     Overlaps with U16153 and U28368.
FEATURES             Location/Qualifiers
     source          1..1147
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p22.3-p23"
                     /cell_line="4251 astrocytoma"
                     /clone_lib="lambda gt10"
     gene            36..521
                     /gene="ID4"
     CDS             36..521
                     /gene="ID4"
                     /codon_start=1
                     /product="helix-loop-helix protein"
                     /protein_id="CAA69255.1"
                     /db_xref="GI:1550770"
                     /db_xref="SWISS-PROT:P47928"
                     /translation="MKAVSPVRPSGRKAPSGGGGGELALRCLAEHGHSLGGSAAAAAA
                     AAAARCKAAEAAADEPALCLQCDMNDCYSRLRRLVPTIIPNKKVSKVEILQHVIDYIL
                     DLQIALETHPALLRQPPPAPPHPAGTCPAAPPRTPLTALNTDPAGAVNKQGDSILC
                     R"
     polyA_signal    1130..1135
     polyA_site      1146
BASE COUNT    280 a  286 c  320 g  261 t
ORIGIN
```

```
Qy    277  GCAAAGGAGACGGGAGCGCGCCGCGCGGGACGGGGTTGCCTGCCTCCCTCGCTGC 336
                  || | |||| || |||||||||||||||||||||||||||||||||
Db   1866  GCGCCGGACCGGCGAGGAGGGGAGCAAAGCTAAAATTGCGGGCGAGCTTCGC 1925

Qy    337  CCCAGCGCGGGT-TCGCTCGGGTAGAGCGCAGGGGGTGAGCGCCGGT 395
           |||||||||||| |||||||||||||||||||||||||||||||||||
Db   1926  TCTACCGGCTTGTCGCGGTTCTCTCCGCAGGAAGCGCGCGGTGAGCCCGGT 1985

Qy    396  GCGCCCCTTCGGGGCGGCGCAAGGCGGGCCGCGGCCTGGGCGTCGGCGTCGGGGTGGGGCTG 455
           |||||||||||||| || |||||||||||||||||||||||||||||||||||||
Db   1986  GCGCCCCTCGGGCGCAAGGCGCGCGGGGGGGGGGGGACTGGGCGTCGGGCCTACCGCTG 2045

Qy    456  CCTGGCCGGAGCACGGCCCACAGCCTGGGTGGGCTCCCGCAGCCCGC 500
                 ||||||||||||||||||||||||||||||||||||||
Db   2046  CCTGGCCGGAGCACGGCCCACCAGCCTGGGTGGGCTCGGGCTGCGGGGGAGCCGCGGGC 2090
```

Query Match 31.7%; Score 158.4; DB 9; Length 1147;
Best Local Similarity 99.4%;
Matches 159; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    341  GCGGGTTCGCTCGCGTAGAGCGCAGGGCGCGGGCGCGGGCGCGGTGAGCGCCGGTGCC 400
```

```
Db    1 GGGGGGTTCGCTCGCGTAGAGCGCAGGGCGCGCCGATGAAGGCGGTGAGCGCCCTGG
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    401 CCTCGGGCGCGGAAGGGCCGCGGTGCGGGCTGGGGGGAGGGTGGGCGTGCGCCTGCCTGG 460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTCGGGCGCGGAAGGGCCGCGGTGCGGGCTGGGGGGAGGGTGGGCGTGCGCCTGCCTGG 120
Qy    461 CCGAGCACGGGCCACAGCCTGGGTGGGCTCCCGAGCACCCGGCGC 500
          |||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGCACGGGCCACAGCCTGGGTGGGCTCCCGAGCACCCGGC 160
```

RESULT 12
LOCUS   AX323848/c
DEFINITION   AX323848   1473 bp   DNA   linear   PAT 07-JAN-2002
             Sequence 336 from Patent WO0192565.
ACCESSION   AX323848
VERSION   AX323848.1 GI:18094597
KEYWORDS   .
SOURCE   synthetic construct.
  ORGANISM   synthetic construct
             artificial sequence.
REFERENCE   1 (sites)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE   Diagnosis of diseases associated with dna transcription
  JOURNAL   Patent: WO 0192565-A 336 06-DEC-2001;
             Epigenomics AG (DE)
FEATURES   Location/Qualifiers
     source   1..1473
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT   347 a   111 c   431 g   584 t
ORIGIN

Query Match   29.9%; Score 149.6; DB 6; Length 1473;
Best Local Similarity   67.1%; Pred. No. 3;4e-15;
Matches   243; Conservative   0; Mismatches 114; Indels   5; Gaps   2;

```
Qy    12 CGGCGGTTGTGAGTAGTAGTACCCGGGAGTGGGGTTGATCCCGGGTCTAGGGGAGCGGCGGGCCCGA 71
         |||||||||||||||||||| | ||||||||| |||||||||||||||| ||||| |||
Db   360 CGGCGGATTATAAAAATAATACCCGGAAAAATAATAATCCGGAACTAAAAAAACGCGACGACCGC 301
Qy    72 TCGGGCTTAGTCGGGAGCTCCGAAGGGAGTGACTAGGACACCCGGGTGGGCTACTTTTCTT 131
         ||| ||||| ||||||||||||||| |||| |||||| || ||||||| |||||| |||||
Db   300 ATCGAACTTAATCGAAACTCCGAAAAAATAACTAAAACACCCGAATAAAACTACTTTTCTT 241
Qy   132 CCGGTGCTTTTGCTTTTTTTTTTCCTTTGGCCTCGGGCTCCCCCACTGAGCAAAG 191
         ||| ||||||||| ||||||||||||| ||||||||||||| || |||||||||||
Db   240 CCGATACTTTTAC-TTTTTTTTTCCTTTAAACTCGAACTAAATATCGCCCACTAAACAAA 182
Qy   192 ATTCCCTCGTAAAAACCAGAGCGACCCTCCCGTCAATTGTTTGGGCTCGGGAGTGTCGGG 251
         |||||| |||| || |||||| || ||||||| |||| |||||||| ||||||| ||||
Db   181 ATTCCCTCGTAAAAACCAAAACGACCCTCCCGTCAATTATTTAAACTCGAAAATATCGGA 122
Qy   252 TGCCCCGAGCGCGCGCGCGGCGGGAGGCAAAGGGAGGCGAGCGGGCGGGGCCG 311
         ||||||||| || |||| || || ||| |||||| || |||||||||||||||
Db   121 TACCCGAACGCGCGCGAAC----CAAACAAAAAAGGAGCGCGACGAAAACCG 66
Qy   312 GAGCTTGCCTGCCTGCCTCCCGCCTCGCGTCGCGTAGAGCGCAGGGCGCG 371
         ||||| || || ||| ||||||||| ||| |||| ||||| |||||||
Db    65 AAACTTACCTACCTCCCGCTCGCCTCGCCCAAGCGAATTCGCTCCGCCCAAAAACGAAAAACGACG 6
Qy   372 CG 373
         ||
Db     5 CG 4
```

RESULT 13
LOCUS   MMID4HLH   1659 bp   mRNA   linear   ROD 09-OCT-1997
DEFINITION   M.musculus mRNA for Id4 helix-loop-helix protein.
ACCESSION   X75018

VERSION   X75018.1 GI:402637
KEYWORDS   helix-loop-helix protein; Id4 protein.
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1659)
  AUTHORS   Sablitzky,F.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-SEP-1993) F. Sablitzky, Max-Delbruck-Labor in der
             MPG, Carl-von-Linne-weg 10, 50829 Koeln, FRG
REFERENCE   2 (bases 1 to 1659)
  AUTHORS   Riechmann,V., van Cruchten,I. and Sablitzky,F.
  TITLE   The expression pattern of Id4, a novel dominant negative
             helix-loop-helix protein, is distinct from Id1, Id2 and Id3
  JOURNAL   Nucleic Acids Res. 22 (5), 749-755 (1994)
  MEDLINE   94188125
FEATURES   Location/Qualifiers
     source   1..1659
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="VR4, B, F and I"
                     /tissue_type="Bone marrow"
                     /clone_lib="lambda gt11 and lambda gt12"
                     /dev_stage="embryo"
     gene   72..557
                     /gene="Id4"
     CDS   72..557
                     /gene="Id4"
                     /codon_start=1
                     /product="Id4"
                     /protein_id="CAA52926.1"
                     /db_xref="GI:402638"
                     /db_xref="MGD:MGI:99414"
                     /db_xref="SWISS-PROT:P41139"
                     /translation="MKAVSPVRPSGRKAPSGCGGGGELALRCLAEHGHSLGSSAAAAA
                     AAAARCKAAEAAAADEPALCLQCDMNDCYSRLRRLVPTIPPNKKVSKVEILQHVIDYIL
                     DLQLALETHPALLRQPPPAPELHPAGACPVAPPRTPLTALNTDPAGAVNKGDSILC
                     R"
BASE COUNT   390 a   373 c   437 g   459 t
ORIGIN

Query Match   26.6%; Score 133.2; DB 10; Length 1659;
Best Local Similarity   91.6%; Pred. No. 1;5e-12;
Matches   141; Conservative   0; Mismatches 13; Indels   0; Gaps   0;

```
Qy   347 TCGCTCGCGTAGAGCGCAGGGCGCGCCGATGAAGGCGGTGAGCGCCCTGG 406
         ||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    43 TCGCGCGTCCTCTCGCCCAGGAAGCGCGGGTGCGGGCGGTGAGCCCCGTGG 102
Qy   407 GCCGCGAAGGGCCGCGGTGCGGGCGGGGAGGGTGGCGCTGCGCCTGGCCGAGC 466
         |||||||||||| | ||||||||||||||||| || |||| |||||||||||
Db   103 GCCCCAAGGCGCGCGTCGGGCGTGCGGCGAGGCGAGCTGGGGGGCTACGCGC 162
Qy   467 ACGGGCCACAGCCTGGGTGGGCTCCCGCAGCCGCGGC 500
         |||||||||| ||||||||||||||||| ||||||||
Db   163 ACGGCCACAGCCTGGGTGGGCTCGGCGAGCCCGCGGC 196
```

RESULT 14
LOCUS   AR103268   1309 bp   DNA   linear   PAT 14-FEB-2001
DEFINITION   Sequence 4 from patent US 6087171.
ACCESSION   AR103268
VERSION   AR103268.1 GI:12814856
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE   1 (bases 1 to 1309)
  AUTHORS   Neuman,T., Suda,K. and Nornes,H.O.
  TITLE   Method for inducing DNA synthesis in neurons

```
RESULT 15

LOCUS       AC097157          195286 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-138P16, *** SEQUENCING IN PROGRESS
            ***, 54 unordered pieces.
ACCESSION   AC097157
VERSION     AC097157.4  GI:17973636
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 195286)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,J., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,P., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Wang,J., Warren,R., Washington,C.,
            Watkins,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195286)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17064510.
            ----------------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------------------- Project Information
            Center project name: GIFA
            Center clone name: CH230-138P16
            ----------------------- Summary Statistics
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
            Consensus quality: 162156 bases at least Q40
            Consensus quality: 169369 bases at least Q30
            Consensus quality: 175407 bases at least Q20
            Estimated insert size: 170489; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation
            ----------------------
            * NOTE: Estimated insert size may differ from sequence length
            *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *  consists of 54 contigs. The true order of the pieces
            *  is not known and their order in this sequence record is
            *  arbitrary. Gaps between the contigs are represented as
            *  runs of N, but the exact sizes of the gaps are unknown.
            *  This record will be updated with the finished sequence
            *  as soon as it is available and the accession number will
            *  be preserved.
            *
            *        1: 8480: contig of 8480 bp in length
            *     8481    8580: gap of unknown length
            *     8581   18958: contig of 10378 bp in length
            *    18959   19058: gap of unknown length
            *    19059   27975: contig of 8917 bp in length
            *    27976   28075: gap of unknown length
            *    28076   34070: contig of 5995 bp in length
            *    34071   34170: gap of unknown length
            *    34171   39835: contig of 5665 bp in length
            *    39836   39935: gap of unknown length
            *    39936   49241: contig of 9306 bp in length
            *    49242   49341: gap of unknown length
            *    49342   56455: contig of 7114 bp in length
            *    56456   56555: gap of unknown length
            *    56556   60831: contig of 4276 bp in length
            *    60832   60931: gap of unknown length
            *    60932   64765: contig of 3834 bp in length
            *    64766   64865: gap of unknown length
            *    64866   72653: contig of 7788 bp in length
            *    72654   72753: gap of unknown length
```

```
ORIGIN
BASE COUNT     305 a   339 c   381 g   284 t
FEATURES             Location/Qualifiers
     source          1..1309
                     /organism="unknown"
JOURNAL     Patent: US 6087171-A 4 11-JUL-2000;

Query Match          23.4%;   Score 116.8;  DB 6;  Length 1309;
Best Local Similarity  61.8%;  Pred. No: 6.8e-10;  DB 6;
Matches  220;  Conservative   0;  Mismatches 132;  Indels   4;  Gaps   2;

Qy       145 TTTTTTTTTCCTTTGGGCTTGAGTGTCGCCCACTGAGCAAAGATTCCCTCGTAA 204
             |||||||| || ||||||| || |||||||| |||||||||||||| |||
Db         1 TTTTTCTTGTCTTTTGTTCTTCGGGCCGATTGTCGCTCACAAAAAAAGAAAAAAAAAA 60

Qy       205 ACCCAGAGCGACCACCTTCCGTCAATTGTTGGGCGTTCGGGGGTGCCCGAGCGCG 264
             |||| | ||| |  ||| ||| ||| ||  |||| |||||| |||||||||||||||
Db        61 ACAAAAAACCTGAGCTACCTTTTCAAAAACCCCGACCAGCCTCTCGTCGATTTCTG---GAG 117

Qy       265 CCCGGCGCGGAGGCAAAAGGGAGCGGCGGCGGCGGCGCGGACGGGGAGCGTTGCCTGCC 324
             |||||||||| ||| ||||  |  |  ||||||||| || ||||||| ||||||||||  |
Db       118 CTCGGAGCGCCGTCGGAGCGCCGGGTTCGGCGCGCGGCAGGAGGAGGGAGCAAAAGCAAAATTGCG 177

Qy       325 TCCCTCGCTCGCCCCAGCGCGGTTGCCTCGGTAGAGCCGCAGGGGCGCGCGGCGGCGATGAGCGCG 384
             |||||| |||||||||| ||| |  || || |  ||| | ||||||||| |||||||| |||||||
Db       178 GCGGGGGACCGAGCTGCGTCGCTCGCTGCTTGTGCGGGCGTCCGGTAGCAGCGGCGCGGATGAAGGCG 237

Qy       385 GTGAGCCCGGTGCGCCCCTCGGGCGCCCCGTGCCGCGTGCGGGGGGGAGCTG 444
             |||||||||||||||||| |||||||||||| |||||||||||||||||||||
Db       238 GTGAGCCCGGTGCG-CCCTCGGGCGTCAAGGCGCCTTGCGCGTGCGGGGGAGCTG 296

Qy       445 GCGGTGCGCGTCTGGCCAGCAGCCACCGCACGCGGGCCCGCCCGTGCCGGGCGCCCGGGGGAGCTG 500
             ||||||||||||||| |||||||||| |||||||| |||||||||||||||||||||||| ||
Db       297 GCGGTGCGGTGCCTCGGCGAGCACAGGCCACAGGGCCACAGCCTGCTGGGGTGGCGTCGGAGCCCGCGGC 352
```

```
72754   78060: contig of 5307 bp in length
78061   78160: gap of unknown length
78161   82082: contig of 3922 bp in length
82083   82182: gap of unknown length
82183   87347: contig of 5165 bp in length
87348   87447: gap of unknown length
87448   91729: contig of 4282 bp in length
91730   91829: gap of unknown length
91830   95216: contig of 3387 bp in length
95217   95316: gap of unknown length
95317   99588: contig of 4272 bp in length
99589   99688: gap of unknown length
99689  105523: contig of 5835 bp in length
105524 105623: gap of unknown length
105624 109024: contig of 3401 bp in length
109025 109124: gap of unknown length
109125 114325: contig of 5201 bp in length
114326 114425: gap of unknown length
114426 117124: contig of 2699 bp in length
117125 117224: gap of unknown length
117225 121879: contig of 4655 bp in length
121880 121979: gap of unknown length
121980 125165: contig of 3186 bp in length
125166 125265: gap of unknown length
125266 128870: contig of 3605 bp in length
128871 128970: gap of unknown length
128971 132289: contig of 3319 bp in length
132290 132389: gap of unknown length
132390 135511: contig of 3122 bp in length
135512 135611: gap of unknown length
135612 139740: contig of 4129 bp in length
139741 139840: gap of unknown length
139841 143299: contig of 3459 bp in length
143300 143399: gap of unknown length
143400 146563: contig of 3164 bp in length
146564 146663: gap of unknown length
146664 150122: contig of 3459 bp in length
150123 150222: gap of unknown length
150223 152094: contig of 1872 bp in length
152095 152194: gap of unknown length
152195 154632: contig of 2438 bp in length
154633 154732: gap of unknown length
154733 157037: contig of 2305 bp in length
157038 157137: gap of unknown length
157138 159016: contig of 1879 bp in length
159017 159116: gap of unknown length
159117 160077: contig of 961 bp in length
160078 160177: gap of unknown length
160178 161077: contig of 900 bp in length
161078 161177: gap of unknown length
161178 162674: contig of 1497 bp in length
162675 162774: gap of unknown length
162775 164327: contig of 1553 bp in length
164328 164427: gap of unknown length
164428 167661: contig of 3234 bp in length
167662 167761: gap of unknown length
167762 169703: contig of 1942 bp in length
169704 169803: gap of unknown length
169804 172362: contig of 2559 bp in length
172363 172462: gap of unknown length
172463 174088: contig of 1626 bp in length
174089 174188: gap of unknown length
174189 175748: contig of 1560 bp in length
175749 175848: gap of unknown length
175849 178001: contig of 2153 bp in length
178002 178101: gap of unknown length
178102 179573: contig of 1472 bp in length
179574 179673: gap of unknown length
179674 181733: contig of 2060 bp in length
181734 181833: gap of unknown length
181834 183356: contig of 1523 bp in length
183357 183456: gap of unknown length
183457 184584: contig of 1128 bp in length
184585 184684: gap of unknown length
184685 186054: contig of 1370 bp in length
186055 186154: gap of unknown length
186155 187367: contig of 1213 bp in length
187368 187467: gap of unknown length
187468 188600: contig of 1133 bp in length
188601 188700: gap of unknown length
188701 189936: contig of 1236 bp in length
189937 190036: gap of unknown length
190037 191328: contig of 1292 bp in length
191329 191428: gap of unknown length
191429 192719: contig of 1291 bp in length
192720 192819: gap of unknown length
192820 193976: contig of 1157 bp in length
193977 194076: gap of unknown length
194077 195286: contig of 1210 bp in length.
```

FEATURES   Location/Qualifiers
  source   1..195286
                  /organism="Rattus norvegicus"
                  /db_xref="taxon:10116"
                  /clone="CH230-138P16"

Query Match 13.3%;   Score 66.4;   DB 2;   Length 195286;
Best Local Similarity   57.0%;   Pred. No. 0.034;
Matches 118; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy    293 CGGCCCGGGACGGGGCGCCCGGAGCTTGCCTGCTCCCCTCGCCCCCCAGCGGGGGTTCGGCTC 352
              | ||||   ||  ||   ||| |  |     ||  ||  |||||||
Db 113699 CCGGCCCGGGCCCCGCCCCCGGCGGGGGGCCCCGCCCCCCGCCCCGCGCCCC 113758

Qy    353 GCGTAGAGCGCAGGGGCGCGCGGCGGATGAAAGGCGCGGTGAGCCCGGTGCGCCCCTCGGGCCGCA 412
          | |   | |||  ||| |||| |  |    |||| ||| ||   | ||| ||||
Db 113759 GCGCCGGGCGCGGGGGCCCGGGGGGGCGNGGGGCGCGNGGCGGGCGCCCGGGGCGCC 113818

Qy    413 AGGGCGCCGTCGGGCGTTCGGGCGGGGGGGAGCTGGCGCTGCCGCCGAGGACGGGC 472
           ||||| | | | || ||  ||||| | ||| ||  ||| || |  |  |  ||||
Db 113819 GAGGGCGGGCGCGCGGGGGGCCCNGGCGGCGGGCGCNGGCGCGCGGGCGGCGGGGGGG 113878

Qy    473 ACAGCCTGGGTGGCTCCGCAGCCGCGG 499
           || || ||| |||| || ||| ||| ||
Db 113879 GCGGGGCGGGGCGCGGGGGGGGGCGCCGGGGG 113905
```

Search completed: October 10, 2002, 17:36:40

Job time : 971.867 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
　　　　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　14616.225 Million cell updates/sec

OM nucleic - nucleic search, using sw model

Title:　　　　　US-09-489-101A-7_COPY_1_500
Perfect score:　500
Sequence:　　　1 cggctcagcggggcggagg...............cgccactatcgccgcggcg 500
Scoring table: IDENTITY_NUC
　　　　　　　　Gapop 10.0 , Gapext 1.0
Searched:　　　1797656 seqs, 1046326293 residues

Total number of hits satisfying chosen parameters:　　　3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　　Maximum Match 100%
　　　　　　　　　Listing first 45 summaries

Database :　　GenEmbl:*
　　　　　　　1: gb_ba:*
　　　　　　　2: gb_htg:*
　　　　　　　3: gb_in:*
　　　　　　　4: gb_Om:*
　　　　　　　5: gb_ov:*
　　　　　　　6: gb_pat:*
　　　　　　　7: gb_ph:*
　　　　　　　8: gb_pl:*
　　　　　　　9: gb_pr:*
　　　　　　　10: gb_ro:*
　　　　　　　11: gb_sts:*
　　　　　　　12: gb_sy:*
　　　　　　　13: gb_un:*
　　　　　　　14: gb_vi:*
　　　　　　　15: em_ba:*
　　　　　　　16: em_fun:*
　　　　　　　17: em_hum:*
　　　　　　　18: em_in:*
　　　　　　　19: em_mu:*
　　　　　　　20: em_Om:*
　　　　　　　21: em_or:*
　　　　　　　22: em_pat:*
　　　　　　　23: em_ph:*
　　　　　　　24: em_pl:*
　　　　　　　25: em_ro:*
　　　　　　　26: em_sts:*
　　　　　　　27: em_un:*
　　　　　　　28: em_vi:*
　　　　　　　29: em_vl:*
　　　　　　　30: em_htg_hum:*
　　　　　　　31: em_htg_inv:*
　　　　　　　32: em_htg_other:*
　　　　　　　33: em_htgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result		Query					
No.	Score	Match	Length	DB	ID		Description

1	500	100.0	2389	6	AX201586	AX201586 Sequence
2	500	100.0	2389	9	HUMMAZ	M94046 Human zinc
3	492.6	98.5	1638	9	HUMZINCFP	M93339 Human zinc
4	492.6	98.5	1767	9	HSU33819	U33819 Human zinc-
5	485.8	97.2	1738	9	D85131	D85131 Homo sapien
6	431.8	86.4	1598	10	L04649	L04649 Mouse mRNA
7	398.2	79.7	9701	9	MUSPUR	AB017335 Homo sapi
8	379	75.8	2535	4	AF076784	AF076784 Oryctolag
9	378.2	75.6	993	10	HAMHIT	L06008 Syrian hams
10	347.2	69.4	5615	10	AB000360	AB000360 Mus muscu
11	138.4	27.7	175599	2	AC009133	AC009133 Homo sapi
12	97.6	19.5	86621	2	AC093933	AC093933 Rattus no
13	96.8	19.4	93821	2	AC021596	AC021596 Homo sapi
14	94.8	18.9	197669	2	AC084064	AC084064 Homo sapi
15	94.6	18.9	137878	2	AP003300	AP003300 Oryza sat
16	94.4	18.8	158167	2	AC034105	AC034105 Homo sapi
17	93.8	18.7	56870	2	AC096161	AC096161 Rattus no
18	93.6	18.7	158167	2	CEY39B6_3	Continuation (4 of
19	93.6	18.7	88064	2	AC034105	AC034105 Homo sapi
20	93.4	18.7	158167	2	AC022707	AC022707 Homo sapi
21	92.8	18.6	300695	2	AC079431	AC079431 Mus muscu
22	92.2	18.4	115715	2	AC079411	AC079411 Homo sapi
23	91.6	18.3	5880	14	HSE4	D14486 Equine herp
24	91	18.2	72360	2	AC022984	AC022984 Homo sapi
25	91	18.2	78220	2	AC023212	AC023212 Homo sapi
26	90.6	18.1	2561	6	AX239605	AX239605 Sequence
27	89.6	17.9	63017	2	AC110291	AC110291 Homo sapi
28	89.2	17.8	840	8	CNS01BNS	CNS01BNS Botrytis
29	89.2	17.8	177883	2	AC046159	AC046159 Homo sapi
30	89	17.8	66114	2	AC091309	AC091309 Mus muscu
31	89	17.8	110737	2	AC011105	AC011105 Homo sapi
c 32	89	17.8	244367	2	AC011105	AC011105 Homo sapi
c 33	88.8	17.8	64998	2	AC074336	AC074336 Mus muscu
c 34	88.8	17.8	76052	2	AC104992	AC104992 Homo sapi
35	88.8	17.8	220469	2	AC023262	AC023262 Homo sapi
c 36	88.4	17.7	46181	2	AC074307	AC074307 Mus muscu
37	88.4	17.7	74138	2	AC105855	AC105855 Rattus no
38	88.4	17.6	75628	2	AC021272	AC021272 Homo sapi
39	87.8	17.6	87283	2	AC021793	AC021793 Homo sapi
40	87.8	17.6	141892	2	AC022856	AC022856 Homo sapi
c 41	87.6	17.5	81417	2	AC023197	AC023197 Mus muscu
42	87.4	17.5	78700	2	AC023265	AC023265 Homo sapi
43	87.4	17.5	87440	2	AC022685	AC022685 Homo sapi
c 44	87.2	17.4	265537	2	AC021226	AC021226 Mus muscu
c 45	87	17.4	98421	2	AC105896	AC105896 Rattus no

ALIGNMENTS

RESULT 1
AX201586
LOCUS　　　　AX201586　　　　　　　　2389 bp　　　DNA　　　linear　　　PAT 30-AUG-2001
DEFINITION　Sequence 7 from Patent WO0153349.
ACCESSION　 AX201586
VERSION　　　AX201586.1　　GI:15391435
KEYWORDS
SOURCE　　　 human.
　ORGANISM　 Homo sapiens
　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　 1　(bases 1 to 2389)
　AUTHORS　　Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
　　　　　　　Chen,Y.T.
　TITLE　　　 Small cell lung cancer associated antigens and uses therefor
　JOURNAL　　Patent: WO 0153349-A 7 26-JUL-2001;
　　　　　　　LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
　　　　　　　SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
　　　　　　　INC. (US)
FEATURES　　　　　　　　　Location/Qualifiers
　　　source　　　　　　　1..2389
　　　　　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　　　　　/db_xref="taxon:9606"

Left column

```
Qy    1  CGGCTCAGCGCGGGGGGCCGGAGGCCCATGTGTTCCCGGTGTTTCCTTGCACGCTCGCCCCC   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  CGGCTCAGCGCGGGGGGCCGGAGGCCCATGTGTTCCCGGTGTTTCCTTGCACGCTCGCCCCC   60

Qy   61  CTTTCCCGTGCTGTTGGGGGTGGGGGGTGGGGGGTGGGGGGTGGGGGGCTGAGCTCCGGTTCTT  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CTTTCCCGTGCTGTGGGCCTGACTCCATGAACTCCTTCCCGCC  120

Qy  121  ACCTTCAGGGTCACGCCCAGAGAACCCCCTGCAGGTCCAGGTCCGGTCCGGCTTTCTT  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  ACCTTCAGGGTCACGCCCAGAGAACCCCCTGCAGGTCCAGGTCGGGGGCTGAGCTCCAGTGCCCGGCTTTCTT  180

Qy  181  TGCCTCCCAGGGCTGCCCCAGAGTCCATTCCAGGTCCGGCGCGCGGCGGCCCCCCACGCC  240
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TGCCTCCCAGGGCTGCCCCAGAGTCCATTCCAGGTCCGGCGCGCGGCGGCCCCCCACGCC  240

Qy  241  CCAGGCCCCCCGGGCGAGCGCCCCCCTCCAGGTGGACTTTGCTCTCCCGGTGCTGGGGGGCCCA  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CCAGGCCCCCCGGGCGAGCGCCCCCCTCCAGGTGGACTTTGCTCTCCCGGTGCTGGGGGGCCCA  300

Qy  301  GGAGTCGCGGCGCGCTGCGTGCTGCCCGCGCCGTCGCCTGCCGCGCCGGC  360
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GGAGTCGCGGCGCGCTGCGTGCTGCCCGCGCCGTCGCCTGCCGCGCCGGC  360

Qy  361  CCCGGCCCCCCGTGCCGCGCGTCTACGGTGGACACAGCAGCAGCGGCCCCGGC  420
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  CCCGGCCCCCCGTGCCGCGCGTCTACGGTGGACACAGCAGCAGCGGCCCCGGC  420

Qy  421  CCCTCCGCGCCACCCCCCCAGTGTCCGGCGGGCGGGCCGGAGGCGCCTCCGGCGCG  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  CCCTCCGCGCCACCCCCCCAGTGTCCGGCGGGCGGGCCGGAGGCGCCTCCGGCGCG  480

Qy  481  CGCCACTATCGCGCCGGCGGGGCGCCG  500
         ||||||||||||||||||||||||||
Db  481  CGCCACTATCGCGCCGGCGGGGCGCCG  500
```

RESULT 2

HUMMAZ

LOCUS   HUMMAZ   2389 bp   mRNA   linear   PRI 27-APR-1993
DEFINITION Human zinc finger protein (MAZ) mRNA.
ACCESSION   M94046
VERSION   M94046.1 GI:187393
KEYWORDS   zinc-finger protein.
SOURCE   Homo sapiens female cDNA to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2389)
AUTHORS   Bossone,S.A., Asselin,C., Patel,A.J. and Marcu,K.B.
TITLE   MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
            regulating transcriptional initiation and termination
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456 (1992)
MEDLINE   92366479
FEATURES   Location/Qualifiers
     source   1..2389
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="HeLa"

Query Match   100.0%; Score 500; DB 9; Length 2389;
Best Local Similarity   100.0%; Pred. No. 1.6e-45;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT   419 a   805 c   731 g   434 t
ORIGIN

Right column

```
Qy    1  CGGCTCAGCGCGGGGGGCCGGAGGCCCATGTGTTCCCGGTGTTTCCTTGCACGCTCGCCCCC   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  CGGCTCAGCGCGGGGGGCCGGAGGCCCATGTGTTCCCGGTGTTTCCTTGCACGCTCGCCCCC   60

Qy   61  CTTTCCCGTGCTGTTGGGGGTGGGGGGTGGGGGGTGGGGGGCTGAGCTCCATGAACTCCTTCCCGCC  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CTTTCCCGTGCTGTTGGGGGTGGGGGGTGGGGGGCTGAGCTCCATGAACTCCTTCCCGCC  120

Qy  121  ACCTTCAGGGTCACGCCCAGAGAACCCCCTGCAGGTCCAGGTCGGGGGCTGAGCTCCAGTGCCCGGCTTTCTT  180
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  121  ACCTTCAGGGTCACGCCCAGAGAACCCCCTGCAGGTCCAGGTCGGGGGCTGAGCTCCAGTGCCCGGCTTTCTT  180

Qy  181  TGCCTCCCAGGGCTGCCCCAGAGTCCATTCCAGGTCCGGCGCGCGGCGGCCCCCCACGCC  240
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TGCCTCCCAGGGCTGCCCCAGAGTCCATTCCAGGTCCGGCGCGCGGCGGCCCCCCACGCC  240

Qy  241  CCAGGCCCCCCGGGCGAGCGCCCCCCTCCAGGTGGACTTGCTCTCCCGGTGCTGGGGGGCCCA  300
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CCAGGCCCCCCGGGCGAGCGCCCCCCTCCAGGTGGACTTGCTCTCCCGGTGCTGGGGGGCCCA  300

Qy  301  GGAGTCGCGGCGCGCTGCGTGCTGCCCGCGCCGTCGCCTGCCGCGCCGGC  360
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GGAGTCGCGGCGCGCTGCGTGCTGCCCGCGCCGTCGCCTGCCGCGCCGGC  360

Qy  361  CCCGGCCCCCCGTGCCGCGCGTCTACGGTGGACACAGCAGCAGCGGCCCCGGC  420
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  CCCGGCCCCCCGTGCCGCGCGTCTACGGTGGACACAGCAGCAGCGGCCCCGGC  420

Qy  421  CCCTCCGCGCCACCCCCCCAGTGTCCGGCGGGCGGGCCGGAGGCGCCTCCGGCGCG  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  CCCTCCGCGCCACCCCCCCAGTGTCCGGCGGGCGGGCCGGAGGCGCCTCCGGCGCG  480

Qy  481  CGCCACTATGCGCGGGGCGGG  500
         |||||||||||||||||||||
Db  481  CGCCACTATGCGCGGGGCGGG  500
```

RESULT 3

HUM2INCFP

LOCUS   HUM2INCFP   1638 bp   mRNA   linear   PRI 24-NOV-1994
DEFINITION Human zinc finger protein mRNA.
ACCESSION   M93339 J05371
VERSION   M93339.1 GI:340439
KEYWORDS   zinc-finger protein.
SOURCE   Homo sapiens (library: lambda gt11) female cDNA to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1638)
AUTHORS   Pyrc,J.J., Moberg,K.H. and Hall,D.J.
TITLE   Isolation of a novel cDNA encoding a zinc-finger protein that binds
            to two sites within the c-myc promoter
JOURNAL   Biochemistry 31 (16), 4102-4110 (1992)
MEDLINE   92232709
FEATURES   Location/Qualifiers
     source   1..1638
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="HeLa"
                     /cell_type="human cervical carcinoma, epithelial."
                     /tissue_lib="lambda gt11"

Query Match   98.5%; Score 492.6; DB 9; Length 1638;
Best Local Similarity   99.2%; Pred. No. 1.1e-44;
Matches 495; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

BASE COUNT   275 a   598 c   515 g   250 t
ORIGIN

```
Qy    2  GGCTCAGCGGGGGGGGCCGAGGCCATGTGTTCCCGGTGTTTCCTTGCACGCTCGCCCC   61
         ||| ||||| ||||||||| | |||||||||||||||||||||||||||||||||
Db   31  GGGCCCGCGTGCGGCGGCCGGAGGCCATGTTCCCGGTGTTTCCGTGCACGCTCGCCCC   90
```

```
RESULT 4
HSU33819
LOCUS       HSU33819      1767 bp    mRNA    linear   PRI 22-JUL-1996
DEFINITION  Human zinc-finger DNA binding protein (MAZ) mRNA, partial cds.
ACCESSION   U33819
VERSION     U33819.1   GI:995934
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1767)
  AUTHORS   Parks,C.L. and Shenk,T.
  TITLE     The serotonin 1a receptor gene contains a TATA-less promoter that
            responds to Maz and Sp1
  JOURNAL   J. Biol. Chem. 271 (8), 4417-4430 (1996)
  MEDLINE   96224025
REFERENCE   2  (bases 1 to 1767)
  AUTHORS   Parks,C.L. and Shenk,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1995) Christopher L. Parks, Molecular Biology,
            Princeton University, Washington Road, Princeton, NJ 08544-1014,
            USA
FEATURES             Location/Qualifiers
     source          1..1767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Hela"
     gene            1..1767
                     /gene="MAZ"
     CDS             <1..1614
                     /gene="MAZ"
                     /note="synonyms: Pur-1, ZF87; zinc-finger DNA-binding
                     protein"
                     /codon_start=1
                     /product="MAZ"
                     /protein_id="AAB04121.1"
                     /db_xref="GI:995935"
                     /translation="EFRLPPAGRGARAGGAGPGRANAFGAAAQPGRPGGAPRARAPRGP
                     RPRPLSPGPAAAEAMFVPFPCTLLAPPFPVIGLDSRGVGGLMNSPPPPQSHAQNPL
                     QVGAELQSRFFASQGCAQSPFQAAPAPPPTPQAPAAEPLQVDLLPVIAAAQESAAAA
                     AAAAAAVAAAAPPAPAAASTVDTAALKQPPAPPPVVSAPAAEAAPPASAATIAA
                     AAATAVVAPPTSTVAVAPVASALEKTKSKGPVICALCAKEFKNGYNLRRHEAIHTGAK
                     AGRVPSGAMKMPTMVPLSLSVPQLSGAGGGGEAGGAGGGAAAVAAGGVTTPASGKR
                     IRKNHACEMCGKAFRDVYLHNRHKLSHSDEKPYQCPVCQQRFKRKDRMSYHVRSHDGA
                     VHKPYNCSHCGKSFSRPDHLNSHVRQVHSTERPFKCEKCEAAFATKDRLRAHVRHEE
                     KVPCHVCGKMMSSAYISDHMKVHSQGPHHVCELCNKGTGEVCPMAAAAAAAAAAA
                     VAAPPTAVGSLSGAEGVPVSSQPLPSQPW"
ORIGIN
BASE COUNT      282 a    654 c    574 g    257 t
```

```
Query Match      98.5%;   Score 492.6;   DB 9;   Length 1767;
Best Local Similarity   99.2%;   Pred No. 1.1e-44;
Matches  495;   Conservative  0;   Mismatches  4;   Indels  0;   Gaps  0;
```

```
Qy     62 TTCCCCGTGTGGGCCTGGACTCCCGGGGGTGGGCGGGGCCTTCATGAACTCCTTCCCGCCA 121
          |||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 TTCCCGTGCTGGGCCTGGACTCCCGGGGGTGGGCGGGGCCTTCATGAACTCCTTCCCGCCA 150

Qy    122 CCTGCAGGGTCACGCCCAGAACCCCTGCAGGTCCAGTCCGAGCTCCAGTCCGGCTTCTTT 181
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    151 CCTCAGGGTCACGCCCAGAACCCCTGCAGGTCCAGTCCGAGCTCCAGTCCGGCTTCTTT 210

Qy    182 GCCTCCCAGGGCTGCGCGCCCAGAGTCCATTCCAGGGCGGCGGGGCGGGGGGCCCCACGCCC 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 GCCTCCCAGGGCTGCGCGCCCAGAGTCCATTCCAGGGCGGCGGGGCGGGGGGCCCCACGCCC 270

Qy    242 CAGGCCCCGGGCGCGGGCGGGCGGGCCGAGCCCCTCCAGGTGGACTTGCTGCCCGGCGCCCCAG 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 CAGGCCCCGGGCGCGGGCGGGCGGGCCGAGCCCCTCCAGGTGGACTTGCTGCCCGGCGCCCCAG 330

Qy    302 GAGTCGCCGCGGCGTGCTGCGGCGCCGGGTGCTCGCTGCCGGCGCCGCCGCCCTCGGGCGCC 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    331 GAGTCGCCGCGGCGTGCTGCGGCGCCGGGTGCTCGCTGCCGGCGCCGCCGCCCTCGGGCGCC 390

Qy    362 CCGGGCGCCCTGCGCGGCGCCCTGCGGGCCCTCGAAGCAGCCGCCTCCGGCGCGCGCCC 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391 CCGGGCGCCCTGCGCGGCGCCCTGCGGGCCCTCGAAGCAGCCGCCTCCGGCGCGCGCCC 450

Qy    422 CCTCCGCGCCACCCCGCGCCGCCCAGTGTCGGGGCGGGCCCCGGGGGCCCCCCGGCGCGCCC 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    451 CCTCCGCGCCACCCCGCGCCGCCCAGTGTCGGGGCGGGCCCCGGGGGCCCCCCGGCGCGCCC 510

Qy    482 GCCACTATCGCCGCGGCGGG 500
          |||||||||||||||||||
Db    511 GCCACTATCGCCGCGGCGGG 529
```

```
RESULT 5
D85131
LOCUS       D85131      1738 bp    mRNA    linear   PRI 06-FEB-1999
DEFINITION  Homo sapiens mRNA for Myc-associated zinc-finger protein of human
            islet, complete cds.
ACCESSION   D85131
VERSION     D85131.1   GI:1752741
KEYWORDS    MAZi; Myc-associated zinc-finger protein of human islet.
SOURCE      Homo sapiens human pancreatic islets cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1738)
  AUTHORS   Tsutsui,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-1996) Hatsumi Tsutsui, RIKEN(The Institute of
            Physical and Chemical Research), Tsukuba Life Science Center;
            3-1-1, Koyadai, Tsukuba, Ibaraki 305, Japan
            (E-mail:tsutsui@rtc.riken.go.jp, Tel:0298-36-3612,
            Fax:0298-36-9120)
```

```
Qy      2 GGCTACGCGGGGCGCTGCTGCGGGCGCCGCCATTGTTCCCGGTGTTTCCGTTCACGCGTGGG 61
          ||| || ||||||| |||||||||||||||||||||||| |||||| ||||||| || ||||
Db    160 GGCCCCCCAGCGGCGGGGCCGCCATGTTCCCGGTGTTTTCGTTCACGCGTGCCCCCCCCC 219

Qy     62 TTCCCCGTGCTGGTGGGCCTGGACTCCCGGGGGTGGGCGGGGCCTTCATGAACTCCTTCCCGCCA 121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 TTCCCCGTGCTGGTGGGCCTGGACTCCCGGGGGTGGGCGGGGCCTTCATGAACTCCTTCCCGCCA 279

Qy    122 CCTCAGGGTCACGCCCAGAACCCCTGCAGGTCGGGGGTTGAGGTCGGGGGCTTGAGCTCCAGTCCG 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 CCTCAGGGTCACGCCCAGAACCCCTGCAGGTCGGGGGTTGAGGTCGGGGGCTTGAGCTCCAGTCCG 339

Qy    182 GCCTCCCAGGGCTGCGCGCCCAGAGTCCATTCCAGGCGCGGCGGGGCGGGGGGCCCCCACGCCC 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 GCCTCCCAGGGCTGCGCGCCCAGAGTCCATTCCAGGCGCGGCGGGGCGGGGGGCCCCCACGCCC 399

Qy    242 CAGGCCCCGGGCGCGGGCGGGCGAGCCCCTCCAGGTGGACTTGCTGCCCGGCGTGCTGCGGCGCCCCAG 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 CAGGCCCCGGGCGCGGGCGGGCGAGCCCCTCCAGGTGGACTTGCTGCCCGGCGTGCTGCGGCGCCCCAG 459

Qy    302 GAGTCGCCGCGGCGCGCTGCTGCGGCGCCGGGTGCTCGCTGCCGGCGCCGCCGCCGCCGTCGGGCGCCCC 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 GAGTCGCCGCGGCGCGCTGCTGCGGCGCCGGGTGCTCGCTGCCGGCGCCGCCGCCGCCGTCGGGCGCCCC 519

Qy    362 CCGGGCCCCGCGCGCGCGCTCTACGGTGGACACAGCGGCGCCCTGAAGCAGCCGCCTCCGGCGCGCCC 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 CCGGGCCCCGCGCGCGCGCTCTACGGTGGACACAGCGGCGCCCTGAAGCAGCCGCCTCCGGCGCGCCC 579

Qy    422 CCTCCGCGCCACCCCGCGCCGTGTCGGGCGGGCCCGGGGGCCCCCCGCGCGCGCCC 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 CCTCCGCGCCACCCCGCGCCGTGTCGGGCGGGCCCGGGGGCCCCCCGCGCGCGCCC 639

Qy    482 GCCACTATCGCCGCGGCGGG 500
          |||||||||||||||||||
Db    640 GCCACTATCGCCGCGGCGGG 658
```

REFERENCE   2 (bases 1 to 1738)
  AUTHORS   Tsutsui,H., Sakatsume,O., Itakura,K. and Yokoyama,K.K.
  TITLE   Members of the MAZ family: a novel cDNA clone for MAZ from human
             pancreatic islet cells
  JOURNAL   Biochem. Biophys. Res. Commun. 226 (3), 801-809 (1996)
  MEDLINE   96428591
FEATURES   Location/Qualifiers
     source   1..1738
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="human pancreatic islets"
     gene   92..1585
                     /gene="MAZi"
     CDS   92..1585
                     /gene="MAZi"
                     /codon_start=1
                     /product="Myc-associated zinc-finger protein of human
                     islet"
                     /protein_id="BAA12728.1"
                     /db_xref="GI:1752742"
                     /translation="MRSRGPARPAAPRPRWSLGPRCAEAMFPVFPCTLLAPPFPVIG
                     LDSRGVGGLMNSRPPPQGHAQNPLQVGAELQSRFEFASQSGCAQSPFQAAAPAPPTPQAP
                     AAEPLQVDLLPVLAAAQSSAAAAAAAAAAAAAAAAAAQAAAAVAAAAPPAPAAASTVDTAALKQPPAPP
                     PPPPVSAPAAEAAPPASAATIAAAAATAVVAPTSTVAVAPVASALEKKTKSKGPYIC
                     ALCAKEFKRGYNIRRHEAIHTGAKAGRVPSGAMKMPTMVPLSLLSVPQLSGAGGGGE
                     AGAGGAAVAAGSVVTTPASGKRIRKNHACEMCKAFRDVYHLNRHKLSHSDEKPYQC
                     PVCQQRRKDRMSYHVRSHDGAVHKPYNCSHCGKSFSRPDHLNSHVRQVHSTERPFK
                     KGTGEVCPMAAAAAAAVAAPPTAVGSLSGAEGVPVSSQPLPSOPW"
BASE COUNT   279 a   641 c   551 g   267 t
ORIGIN

Query Match   97.2%; Score 485.8; DB 9; Length 1738;
Best Local Similarity   99.6%; Pred. No. 5.8e-44;
Matches 487; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   12 GGGCCGAGGGCCATGTTCCCGGTGTTTCCTTGCACGCGTGGCCCCCCCCCCCCTTCCCCGTGC 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   159 GCGCCGAGGCCATGTTCCCGGTGTTTCCTTGCACGCGTGGCCCCCCCCCCCCTTCCCGTGC 218

Qy   72 TGGGCCTGGACTCCCGGGGGGTGGGGCGGCCTCATGAACTCCTTCCCGCCACCTCAGGGTC 131
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   219 TGGGCCTGGACTCCCGGGGGGTGGGGCGGCCTCATGAACTCCTTCCCGCCACCTCAGGGTC 278

Qy   132 ACGCCCAGAACCCCCTGCCAGGGTCGGGGGCTGGAGCTCCAGTCCCGCCGCTTCTTTGCCTCTCCCAGG 191
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   279 ACGCCCAGAACCCCCTGCCAGGGTCGGGGGCTGGAGCTCCAGTCCCGCCGCTTCTTTGCCTCTCCCAGG 338

Qy   192 GCTGCGCCCCAGAGTCCATTCCAGGGAGACTTGCTCCCGGGGTCCCGGGTGCTGCTCGCGCCGAGGCCCCGG 251
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   339 GCTGCGCCCCAGAGTCCATTCCAGGGCGCGCCGCCCCGCCCCACGCAGCGCCCAGGGTCCGCG 398

Qy   252 CGGCCCAGAGCCCCTCCAGGTGGACTTGCTCGCGCCGGTGCTGCGCGGGGTCCAGGAGTCGCGG 311
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   399 CGGCCGAGCCCCTCCAGTGGACTTGCTCGCGCCGGTGCTGCGCGGGGTCCAGGAGTCGCGG 458

Qy   312 CGGCTGCGCGGCCCCAGCCCCTCCAGGTGGACTTGCTCGCGCCGGTGCTGCGCGGGGTCCGCG 371
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   459 CGGCTGCGCGGCCCCTCCAGGTGGACAGCAGCGCCCTCCGGCGCGCCCCCCCCCCCTG 518

Qy   372 CCGCGCCCTCTACGGTTGGACACAGCGCCCTCCGGCGCGCCCCCTCCGCCAC 431
         |||||||||||||||||||||||||||||||||||
Db   519 CCGCGCCCCTCTACGGTTGGACACAGCGCCCTCCGGCGCGCCCCCTCCGCCAC 578

Qy   432 CCCCCCAGTGTCGCGCGCCGCGAGGCCGCCCCCCCGCCGCCGCCGCCACTATCG 491
         |||||||||||||||||||||||||||||||||||||||||||||
Db   579 CCCCGCCAGTGTCGGCGCCCCGGCCGCCGCCGCCTCCGCGCCGCCACTATCG 638

Qy   492 CCGGCGGCGG 500
         ||||||||||
Db   639 CCGGCGGCGG 647

RESULT 6
LOCUS   MUSPUR   1598 bp   mRNA   linear   ROD 27-APR-1993
DEFINITION Mouse mRNA sequence.
ACCESSION   L04649
VERSION   L04649.1 GI:200590
KEYWORDS   .
SOURCE   Mus musculus pancreatic tumor cDNA to mRNA.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1598)
  AUTHORS   Kennedy,G.C. and Rutter,W.J.
  TITLE   Pur-1, a zinc-finger protein which binds to purine-rich sequences,
            activates an insulin promoter in heterologous cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502 (1992)
  MEDLINE   93087555
COMMENT   *cell line = bTC3.
FEATURES   Location/Qualifiers
     source   1..1598
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="bTC3"
                     /tissue_type="pancreatic tumor"
BASE COUNT   287 a   566 c   487 g   258 t
ORIGIN

Query Match   86.4%; Score 431.8; DB 10; Length 1598;
Best Local Similarity   91.6%; Pred. No. 3.7e-38;
Matches 457; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy   2 GGGCTCAGCGCGGGGCGGGCGGAGGCCCAТGTTCCCGGTGTGTTTCCTTTGCACGCGTGGCCCCCCCC 61
         | |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13 GCCCCCACTGTGGCCGGAGCCATGTTCCCGGTGTTTCCTTGCACGCGTGGCCCCCCC 72

Qy   62 TTCCCCGTGCTGCTGGGCCTGGACTCCCGGGGGGTGTTGCGCGGCCTCATGAACTCCTTCCCGCCA 121
         ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   73 TTCCCGGTGCTGGGCCTGGACTCCCGGGGGGGGGGGGGGTGGGCGGCCTCATGAACTCCTTCCCGCCA 132

Qy   122 CCTCAGGGTCACGCCCCAGAAAACCCCTGCCAGGTCGGGGGCTGGAGCTCCAGTCCCGCCGCTTCTTT 181
         |||||||||| ||||||||||||||||||||||||||| |||||||||||||||||||||
Db   133 CCTCAGGGTCACGGCCCAGAGAACCCCCTGCCAGGTCGGGGGCTGAGCTCCAGTCCCGCCGCTTCTTT 192

Qy   182 GCCTCTCCCAGGGGCTGCGCCCCAGAGTCCATTCCAGGCGCGCCGCCCCGCCCCACGCAGCGCCC 241
         |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db   193 GCCTCTCCCAGGGCTGCGCCCCAGAGTCCATTCCAGCCGCGCCGCCCCGCCCCACGCAGCGCCC 252

Qy   242 CAGGCCCCGGCGGCGGGGCCCCTCCAGGTGGACTTGCTCCCGGGGTGCTGCGCGCGGGCCAG 301
         |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   253 CAGGCCCCGGACTTGCTCGCGCCCCTCCAAGTGGACTTGCTCGCGCCGGTTCTCGCGCCGCAG 312

Qy   302 GAGTCGCGGCGGCCGGGGCGGCGTGCTGCGCGGCCGCGCGCCCTGAAGCAGCCTCCGCGCCC 361
         |||||||||| ||||||||||||||||| |||||||||||||||||||||||||||||||
Db   313 GAATCACGCGCCAGGCGCGCGGAGGCCCCTCCAAGTGGACAGCAGCGCCCTCCGCGCGGGTTACTCACCC 372

Qy   362 CCGGCCCCCTGCCGCGCGCGCGCTCTACGGTGGACACACGGCCCTCCGCGCGGTCCGCGCGCGGCGCCCGCGCCC 421
         ||||| ||||| |||||||| ||||||||||||||||||||||||||||||||||||||||
Db   373 CCGGCCCGTGCCGCGCGCCTTCACAGTGGACACACGGCCCTCCGCGCGCCCGAGAGCAGCGCCTCCGCGCCGCCCT 432

Qy   422 CCTCGCGCCACCCCCCGCGCCAGTGTGGGCGCCGCCCGCGCGGCGGCGGCCGCCCTCCGCGCC 481
         || ||| |||||||||| |||| ||||| |||||||||||||||||||||
Db   433 CCTCCGCGCCACCCCCTGCCGCGTCTCCGCACCAGCGCCCGAGGCGCCGCGCGCCCCCC 492

Qy   482 GCCACTATCGCGCGGGGCGG 500
         ||||||||||||| || | |
Db   493 GCCACTATCGCAGCAGGGG 511

```
ACCESSION   AB017335 D89880
VERSION     AB017335.1  GI:3582440
KEYWORDS    Myc-associated zinc finger protein; kinesin-like DNA binding
            protein.
SEGMENT     3 of 3
SOURCE      Homo sapiens lymphocytoblastoma DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Song,J., Murakami,H., Tsutsui,H., Tang,X., Matsumura,M.,
            Itakura,K., Kanazawa,I., Sun,K. and Yokoyama,K.K.
  TITLE     Genomic organization and expression of a human gene for
            Myc-associated zinc finger protein (MAZ)
  JOURNAL   J Biol. Chem. 273 (32), 20603-20614 (1998)
  MEDLINE   98351105
REFERENCE   2  (sites)
  AUTHORS   Song,J., Murakami,H., Yang,Z.Q., Koga,C., Adati,N., Murata,T.,
            Geltinger,C., Saito-Ohara,F., Ikeuchi,T., Matsumura,M., Itakura,K.,
            Kanazawa,I., Sun,K. and Yokoyama,K.K.
  TITLE     Human genes for KNSL4 and MAZ are located close to one another on
            chromosome 16p11.2
  JOURNAL   Genomics 52 (3), 374-377 (1998)
  MEDLINE   99009323
REFERENCE   3  (bases 1 to 9701)
  AUTHORS   Song,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1998) Jun Song, The Institute of Physical and
            Chemical Research (RIKEN), DNA Bank; Koyadai, 3-1-1, Tsukuba,
            Ibaraki 305-0074, Japan (E-mail:song@rtc.riken.go.jp,
            Tel:81-298-36-3612, Fax:81-298-36-9120).
COMMENT     D89880: submitted (16-Dec-1996).
FEATURES             Location/Qualifiers
     source          1..9701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="lymphocytoblastoma"
     gene            join(AB017333.1:96..165,AB017334.1:73..3222,1081..3621)
                     /gene="KNSL4"
     CDS             join(AB017333.1:1..165,AB017334.1:73..3222,1081..3621)
                     /gene="KNSL4"
                     AB017334.1:2801..2954,AB017334.1:3086..3222,1081..1248,
                     1760..1919,2309..2376,AB017334.1:3485,3574..3621)
                     AB017334.1:2148..2357,AB017334.1:2437..2667,
                     AB017334.1:1547..1677,AB017334.1:1769..1920,
                     /gene="KNSL4"
                     /codon_start=1
                     /product="kinesin-like DNA binding protein"
                     /protein_id="BAA33063.1"
                     /db_xref="GI:3582442"
                     /translation="MAAGGSTQQRREMAASAAAISAGRCRLSKIGATRRPPPARVR
                     VAVRLRPFVDGTAGASDPCVRGMDSCSLEIANWRNHQETLKYQFDAFYGERSTQQDI
                     YAGSVQPILRHLEGQNAKVAVLAYGPTGAGKTTHAGQPRATWGDPASGHGPPAAHKGG
                     GCRGPAMGLSVTMSYLEIYQEKVLDLLDPASGDLVIREDCRGNILIPGLSQKPISSFA
                     DFERHFLPASRNRIVGAITRLNQRSRSHAVLLVKVDQRERLAPFQREGKLILIDLAG
                     SEDNRTGRNGLRLKESGAINTSLFALGKVVDALNQGLPRVPYRDSKLTRLLQDSLGG
                     SAHSTIANIAPERRYLDTVSALNFAARSKEVINRPFTNESLQPHALGPVKLSQKEL
                     LGPPEAKRARGPEEEIGSPEPMASSSLCLPETQPPTEAKAAWTRPCGAPPQLGPSAC
                     LPGPEGAPLLSTPKRERMVLMKTVEEKDLEIERLKETKEELEAKMLAQKAEEKRTIVP
                     QCSGPLSHRIVTGAKPLKKAVVMPLQLIQEQOAASPNAEIHILKNKGRKRKLESLDALE
                     PEERAEDCWELOISPELLAHGROKILDLINEGSARDLRSLQRIGPKRAQLIVGWRELH
                     GPFSQVEDLERVEGITGKOMESFLKANILGLAAGORGAS"
     exon            1081..1248
                     /gene="KNSL4"
                     /number=8
     exon            1760..1919
                     /gene="KNSL4"
                     /number=9
     exon            2309..2376
                     /gene="KNSL4"
                     /number=10
     exon            3125..3337
                     /gene="KNSL4"
                     /number=11
     exon            3426..3485
                     /gene="KNSL4"
                     /number=12
     gene            3574..>3621
                     /gene="KNSL4"
                     /number=13
     CDS             5012..8544
                     /gene="MAZ"
                     /number=1
     exon            <5012..5204
                     /gene="MAZ"
                     /number=1
     gene            5012..8544
                     /gene="MAZ"
     CDS             join(5013..5204,5289..6139,6543..6606,6883..7054,
                     8390..8544)
                     /gene="MAZ"
                     /codon_start=1
                     /product="Myc-associated zinc finger protein"
                     /protein_id="BAA33064.1"
                     /db_xref="GI:3582443"
                     /translation="MFPVFPCTLLAPPFPVGLDSRGVGGLMNSFPPPQGHAQNPLQV
                     GAELQSRFFASQGCAQSPFQAAPAPTPAAPEPLQVDLLPVLAAAQESAAAAAAA
                     AAAAAVAVAAPPAPAAAASTVDTAALKQPPAPAPPPPPPVSAPAAAEAAPPASAATIAAAA
                     ATAVVAPTSTVAVAPVASALEKKTKSKGPYICALCAKEFKNGYNLRRHEAIHTGAKAG
                     RVPSGAMKMPTMVPLSLLSVPQLSGAGGGGGEAGAGGGGAAAVAAGGVVTTASGKRIR
                     KNHACEMCGKAFRDVYHLNRHKLSHSDEKPYQCPVCGVCQORFKRKDRMSYHVRSHDGAVH
                     KPYNCSHCGKSFSRPDHLNSHVRQVHSTERPFKCEKCEAAFATKDRLRAHTVRHEEKV
                     PCHVCGKMLSSAYISDHMKVHSQGPHHVCELCNKGTGEVCPMAAAAAAAAAAAAAVA
                     APPTAVGSLGAEGVPVSSQPLPSOPW"
     exon            5289..6139
                     /gene="MAZ"
                     /number=2
     exon            6542..6606
                     /gene="MAZ"
                     /number=3
     exon            6883..7054
                     /gene="MAZ"
                     /number=4
     exon            8390..>8544
                     /gene="MAZ"
                     /number=5
BASE COUNT     1885 a     3013 c     2946 g     1857 t
ORIGIN
```

```
Query Match       79.7%;  Score 398.6;  DB 9;  Length 9701;
Best Local Similarity   84.9%;  Pred. No. 7.5e-35;
Matches  495;  Conservative   0;  Mismatches   4;  Indels   84;  Gaps   1;

Qy        2 GGCTCAGGGGGGGGCCGGAGGCCATGTTCCCGGTGTGTTTCCTTGCACGCTGCTGGCCCCC    61
            |||| ||||||||||  || ||||||| ||||||||  ||| |||||||||||||||||||
Db     4992 GGCCCCCCTGCGGCCCGGAGGCCATGTTCCCGGTGTGTTTCCTTGCACGCTGCTGGCCCCC  5051

Qy       62 TTCCCCGTGCTGGCCCTGGACTCCCGGGGGGTTGGGCGGGCCCTCATGAAACTCCTTCCCGCA    121
            ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     5052 TTCCCCGTGCTGGCCCTGGACTCCCGGGGGGTTGGGCGGGCCCTCATGAAACTCCTTCCCGCA  5111

Qy      122 CCTCAGGGTCACGCCCAGAACCCCCTGCAGGTCGGGCGTCGGAGTCCCAGGTCCCGGCTTCTTT    181
            |||| ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db     5112 CCTCAGGTGTCACGCCCAGAAACCCCTGCAGGTCGGGCGTCGGAGTCCCAGGTCCCGGCTTCTTT  5171

Qy      182 GCCTCCCAGGGCTGCGCCCCAGAGTCCATTC----------------------    211
            |||||||||||||||||||||||||||||                        |||||
Db     5172 GCCTCCCAGGGCTGCGCCCCAGAGTCCATTCCAGGTGAGTGAGTAGGGCGGGCGCGGGGGCGGGCCG  5231

Qy      212 ----------------------------------------CAGGCC    217
                                                    ||||||
Db     5232 GGCTGGGGGGGGGGGGAGCGGCCGGCGCCGGGCCCACGCCCGGGCCCGGCCTTGTGTCTCCGGCAGGGC  5291

Qy      218 GCGCGGGGGGGGGGCCGCCCCCCGGCCCCACGCGCCCCAGGGCCCCGCGGCCGCCGCCCTTCCAGGTGGACTTG    277
            |||||||| |||||||||||||||||||||||||||||||||||||||||| |||||| |||||||||||||
Db     5292 GCGCGGGGGGGGGCCGCCCCCCGGCCCCACGCGCCCCAGGGCCCCGCGGCCGCCGCCCTTCCAGGTGGACTTG  5351

Qy      278 CTCCCGGTGGCTCGCGCGCGCCCCAGGAGTCCGCCGCGCTGCGCGCGCCGCGCCGCT    337
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        GCCGCCGGCGCGCGCGCACGCGCCCCGGCCCCTTCCAGGTGGACTTG  5351
```

```
Db   5352 CTCCCGGTGCTGCTCGCGCGCCCAGGAGTCGCGCGGCGGGCTGCTGCGGCCGGCT
Qy    338 GCTGCCCGCGTCCGCTGCCGGCGCGCCGCCTCCTTACGGTGGACACAGCG 397
          ||||||||||| || |||| |||||||| ||||||||||||||
Db   5412 GCTGCCGCGTCGCTCGCGCGCGCGCGCGCCCCGCCGCGCCTCTACGGTGGACACAGCG 5471
Qy    398 GCCCTGAAGCAGCCTCCGGCGCCCCCCACCCGCGCCCAGTGTCGCGCCGCGCC 457
          |||||||||||||||||||||| |||||| ||||||||||||||||||||||||
Db   5472 GCCCTGAAGCAGCCTCCGGCGCGCGCCCCCCCACCCCGCGCCCAGTGTCGCGCCGCGCC 5531
Qy    458 GAGGCGCGCGCGCGCGCGCGCGCGCGCCCCCGCGCCCCAGTGTCGGCGCCCCGGCCG 500
          ||||||||||||| || ||||| |||||| ||||||| ||||||||||||
Db   5532 GAGGCGCGCGCGCCGCCGCCGCTCCGCGCGCCACTCATCGCGCGGGCGCGG 5574
```

```
RESULT 8
LOCUS       AF076784              2535 bp    mRNA    linear   MAM 09-DEC-1998
DEFINITION  Oryctolagus cuniculus serum amyloid A-activating factor SAF-1 mRNA,
            complete cds.
ACCESSION   AF076784
VERSION     AF076784.1  GI:3986437
KEYWORDS    .
SOURCE      rabbit.
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 2535)
  AUTHORS   Ray,A. and Ray,B.K.
  TITLE     Isolation and functional characterization of cDNA of serum amyloid
            A-activating factor that binds to the serum amyloid A promoter
  JOURNAL   Mol. Cell. Biol. 18 (12), 7327-7335 (1998)
  MEDLINE   99038237
REFERENCE   2  (bases 1 to 2535)
  AUTHORS   Ray,B.K. and Ray,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1998) Veterinary Pathobiology, University of
            Missouri, 313 Connaway Hall, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1..2535
                     /organism="Oryctolagus cuniculus"
                     /db_xref="taxon:9986"
     CDS             14..1447
                     /codon_start=1
                     /product="serum amyloid A-activating factor SAF-1"
                     /protein_id="AAC84007.1"
                     /db_xref="GI:3986438"
                     /translation="MPPVFPCTLLAPPFPVIGLESRGRPSNWSSFIFQGHAQNPLQVG
                     AELQSRFFASQGCAQSPFQAAPAPPPTPQRPAAEPPQVDLLPVFAAAQESAAAAAAAA
                     AAAAVVTAPPAPAASIVDTAALKQPPAPPPPPPEVPSAPAAEAAPPASAATIAAAA
                     AATAVVAPTSTVAVAPVAASALEKRKRSKGPYICALCAKEFKNGYNLRRHEAIHTGAKA
                     GRVPSGAMKMPTMVPLSLLSVPQLSGAGGGEAGAGGGRAAAVAACGVVTTASGKRI
                     RKNHACEMCGKAFRDVYHLNRHKLSHSDERPYQCPVCQQRFKRDRMSYHVRSHDGAV
                     HKPYNCSHCGKRSRPDHLNSHVRQVHSTERPFKCEKCEAAFATKDRLRAHTVRHEEK
                     VPCHVCGKMLSSAYISDHMKVHSQSPHHVCEICNKGTGEVCPMAAAAAAAAAAAAAA
                     APPTAVGSLSGAEGVPVSSQPLPSQPW"
BASE COUNT      411 a    879 c    825 g    420 t
ORIGIN
```

```
Query Match      75.8%;  Score 379;  DB 4;  Length 2535;
Best Local Similarity   88.6%;  Pred. No. 1.5e-32;
Matches  435;  Conservative   0;  Mismatches  50;  Indels   6;  Gaps   2;

Qy     13 GGCGGAGGGCCATGTTCCCGGTGTTGTTTCCTTGCACGGTGCTGCGGCCCCCGCCCGTGCT 72
          |||||||||||||||||||| | ||||| |||||||| ||||||||||| |||||
Db      4 GGCCGAGGCCATGTTCCCAGTGTTCCCTTGCACGTCGCCCCCCCTTCCCCCCGTGCT 63
Qy     73 GGGGGCCTGGACTCCCGGGGGCGGGCGGGCGGCTCTCAGGCCTCAGGGTCA 132
          |||||||||||||| |||||||||||||||||||
Db     64 GGGCTTGGAATCCCGGGCGCCCAGCAACTGGAGCAGCTTC---ATCTTCAGGGTCA 120
Qy    133 CGCCCAGAACCCCTGCAGGTCGGGCGGGGGCGGGG 192
          |||||||||||||||||||||||||||||||||||||
Db    121 ...
```

```
Db    121 CGCCCAGGAACCCCTGCAGGTCGGGGGGCTGAGCTCCAGTCCGGTTGCTTTTGCCTCCCAGGG 180
Qy    193 CTCGCGCCCAGAAGTCCATTCCAGGCCAGCGCCCGCCAGGGCCGCCCCGGC 252
          |||| |||||||||| |||||||| |||| ||||||||||| |||||
Db    181 CTTCGCGCCCAGAGTCCATTCCAGGCGCGCGCCCCCCAGAGCCGCCCCGGC 240
Qy    253 GGGCGGAGGGCCCTCCAGGTGGACTTGCTCCCGGTGCTGCGCGGCCCCGCGC 312
          ||||||||||||||||||||| ||||||||||||||||||||||||
Db    241 GGGCGAGGCCCTTCCAGGTGGACTTGCTCCCGGTGCTGCGGCCCCGGGCGC 300
Qy    313 GGGTGCTCGGCGGCCGCGTGCCGCCCGCTGCTCGGCGCTG---CCGCGCGCCCCCC 369
          ||| || |||||||||| ||| |||| |||||| |||||||   ||||||||
Db    301 GGCTGCTCGGCGCCGGTGCCGCCCTGGTGACGGGCGCGGGCCCGCCCAGCCCCCC 360
Qy    370 TGCCGCGCCCCTTCTACGGTGGACACAGCGGCCCTGAAGCAGCAGCCTCCGGCCGCGCC 429
          ||||||||| |||||||| |||||||||||||| || |||||||| |||||| ||||||
Db    361 CGCCGCCCGCGTCCACTGTGGACACAGCGGCCCCTGAAGCAGCAGCCCCCGGCCCACTAT 420
Qy    430 ACCCCCCCGCCAGTGTCGGGCGCCCGCGGGCCCGCCGAGCCTCGGCGCGCCACCAT 489
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    421 GCCCCCCCGCCCGTGTCGGGCCCCCCGCCCGCGCGCCCGCCGAGCCTCGGCGCGCCACCAT 480
Qy    490 CGCCGCGGCGG 500
          |||| |||
Db    481 CGCCGCGCGG 491
```

```
RESULT 9
LOCUS       HAMHIT                 993 bp    mRNA    linear   ROD 27-APR-1993
DEFINITION  Syrian hamster mRNA sequence.
ACCESSION   L06008
VERSION     L06008.1  GI:191369
KEYWORDS    .
SOURCE      Mesocricetus auratus
  ORGANISM  Mesocricetus auratus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Mesocricetus.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   Kennedy,G.C. and Rutter,W.J.
  TITLE     Pur-1, a zinc-finger protein which binds to purine-rich sequences,
            activates an insulin promoter in heterologous cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502 (1992)
  MEDLINE   93087555
COMMENT     *cell line = HIT (insulinoma).
FEATURES             Location/Qualifiers
     source          1..993
                     /organism="Mesocricetus auratus"
                     /db_xref="taxon:10036"
                     /cell_line="HIT (insulinoma)"
                     /tissue_type="pancreas (endocrine)"
BASE COUNT      168 a    377 c    303 g    145 t
ORIGIN
```

```
Query Match      75.6%;  Score 378.2;  DB 10;  Length 993;
Best Local Similarity   91.38%;  Pred. No. 2.4e-32;
Matches  401;  Conservative   0;  Mismatches  38;  Indels   0;  Gaps   0;

Qy     62 TTCCCCGTGCTGTGGGCCTGGACTCCCGGGGGGGGGGGGTGGGGGCGGCCTCATGAACTCCTTCCCGCCA 121
          |||| ||||||| |||||| |||||||||||||||||||||||||||||||||||||| |||||| |||
Db      1 TTCCCCGGTGCTGTGGGCCTGGACTCCCGGGGGGGGGGGTGGGGGCGGCCTCATGAACTCCTTCCCCGCCA 60
Qy    122 CCTCAGGGTCACGCGCCCAGAACCCCCTGCAGGTCGGGCGGGCCGGGGGGTGGGGGCCGGCCCTTCTTT 181
          ||||||||||||||||||||| |||||||||||||||||||||| ||||||||||||| |||||||
Db     61 CCTCAGGGTCACGGCCCCAGAACCCCCTGCAGGTCGGGCGGGGCTGGGGGTGGGGGCCGGCGCTTCTTT 120
Qy    182 GCCTCCCAGGGCTGCGGGCCCAGAGTCCATTCCAGGCGCGCGCCCCCCAGGGCCGCCCCACCGCC 241
          ||| |||||||| |||||||| |||||||| |||||||| |||||||||||||||||||| |||||
Db    121 GCCTCCCAGGGCTGGGCCCCAGAGTCCATTCCAGGCGCGCCCCCCCAGGGCCGCCCCACCGCC 180
Qy    242 CAGGCCCCGGCGGGGCGGGGCCCAGCCCCTCCAGGTGGACTTGCTCCCGGTGCTGCGGCCCCAG 301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ...
```

```
Qy    181  CAGGCCCCGGCGGGGGGGGCGGGGAGCCCTCCAAGTGGACTTGCTGCTTCGCCGGTTCTCGCCGCCCGCAG  240
           ||||||||| ||||||||||||||||| |||||| ||||| || || ||||||||||||| ||||||||||
Db    302  GAGTGCCGGCGGCGGCGGCTGCGGCGGCGGGCCGGGCTGCGGGCCGCCCGCTGCGGCTGCTCGCCGCCCCC  361

Qy    241  GAATCGGCGCGCGGCGGGGGGCGGGCGCTGCGGGCCGCCCGCTGCAGTGGTTACTGCCCC  300
           |||||||||| |||||||||||||||||||| |||||||||||||||||| ||||||||
Db    362  CCCGGCCCCTGCGCGGCCGCCCGTTCCGTGGTGGACACAGCACAGCGCCTCCGGGCCCCT  421

Qy    301  CCAGCCCCTGCTGCGCGCCCTCCACACAGTGGCACACAGCGCCGGCGGCCCTGAACGACCC  360
           |||||||| ||||||||||| | ||||||||||||||||||||| ||||| |||||||||
Db    422  CCTCCGCCACCCCCAGTGTGTCGGGCCCGGGCCCGGGCGGGCCCCGGCGGCCCCTCGCGC  481

Qy    361  CCTCCGCCACCCCCTGCGCGGTATCCGCGCGGCCCCCGGCGGCCGCCGCCGCGCGCGCGGG  420
           |||||||||||||  |||  ||| ||||||||||||||| ||| ||  || |  ||
Db    482  GCCACTATGCGCGCGCGG                                          500

Qy    421  GCCACCATCGCCGCCAGCGCG  439
           |||||||||| ||||||| |||
Db    482  GCCACTATGCGCGCGCGG  500
```

```
RESULT 10
LOCUS          AB006360          5615 bp          DNA          linear          ROD 14-APR-2000
DEFINITION     Mus musculus DNA for MUSPUR, partial cds.
ACCESSION      AB006360
VERSION        AB006360.1          GI:4521187
KEYWORDS       MAZ; MUSPUR; MYC-associated zinc-finger protein.
SOURCE         Mus musculus (strain:129) female liver DNA.
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (sites)
AUTHORS        Song,J., Murakami,H., Tsutsui,H., Ugai,H., Geltinger,C., Murata,T.,
               Matsumura,M., Itakura,K., Kanazawa,I., Sun,K. and Yokoyama,K.K.
TITLE          Structural organization and expression of the mouse gene for Pur-1,
               a highly conserved homolog of the human MAZ gene
JOURNAL        Eur. J. Biochem. 259 (3), 676-683 (1999)
MEDLINE        99192812
REFERENCE      2 (bases 1 to 5615)
AUTHORS        Murakami,H.
TITLE          Direct Submission
JOURNAL        Submitted (06-AUG-1997) Hiroo Murakami, Life Science Center, RIKEN,
               DNA Bank; 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan
               (E-mail:murakami@rtc.riken.go.jp, Tel:0298-36-3612,
               Fax:0298-36-9120)
COMMENT        Sequence updated (13-Mar-1998).
FEATURES             Location/Qualifiers
     source          1..5615
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /sex="female"
                     /tissue_type="liver"
     gene            join(1804..1995,2066..2920,3282..3347,3608..3781,
                     5057..5209)
                     /gene="MAZ"
     exon            1804..1995
                     /gene="MAZ"
                     /number=1
     CDS             join(<1804..1995,2066..2920,3282..3347,3608..3781,
                     5057..5209)
                     /gene="MAZ"
                     /codon_start=1
                     /product="MUSPUR"
                     /note="MYC associated zinc-finger protein"
                     /function="transcriptional zinc-finger protein"
                     /protein_id="BAA76280.1"
                     /db_xref="GI:4521188"
                     /translation="MFPVFPCILLAPPPVLGLDSRGVGGILMNSFPPPGGHAQNPLQV
                     GAELQSRFFASQGCAQSPFQQARPAPAPPTPQAPAEAEPLQVDLLPVLAAAQESAAAAA
                     AAAAAAAVVTAPPAPAAAASTVDTAALKQPAPPAPAYSAPAEAEPAAEPAAAATIAAA
                     AATAVAVAPTSTAVAVAPVASVLEKKTKSGPYICALCAKEFKNGYNLRRHEAIHTGAKA
                     GRVPSGAMKMPTMVPLSLLSVPQLSGASGGGEAGAGGGTTAVAAGGVVTTTASGKRI
                     RKNHACEMCGKAFRDVYHLNRHKLSHSDEKPYQCPVCQQRFKRKDRMSYHVRSHDGAV
```

```
BASE COUNT    1016 a   1914 c   1640 g   1045 t
ORIGIN
```

```
Qy    HKPYNCSHCGKSFSRRPDHLNSHVRQVHSTERPFKCEKCEAAFATKDRLRAHTVRHEE
      KVPCHVCGKMLSSAYISDHHKVHSQGPHHVCELCNKGTGEVCPMAAAAAAAAAAAAV
      VAAPPPAVGSLSGAEGVPVSSQPLPSQPW"
```

```
RESULT 11
LOCUS          AC009133          175599 bp          DNA          linear          HTG 25-APR-2001
DEFINITION     Homo sapiens chromosome 16 clone RP11-504I2, WORKING DRAFT
               SEQUENCE, 5 ordered pieces.
ACCESSION      AC009133
VERSION        AC009133.6          GI:13786313
KEYWORDS       HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE         Homo sapiens (human).
ORGANISM       Homo sapiens
               human.
```

```
Query Match        69.4%;   Score 347.2;   DB 10;   Length 5615;
Best Local Similarity   79.7%;   Pred. No. 2.9e-29;
Matches  456;   Conservative   0;   Mismatches  43;   Indels  73;   Gaps  1;

Qy      2  GGCTCACGGGGGGCCGAGGCCCGGCCCCATTGTTCCCGGTGTTTCTTGCACGGTCGTTGTGGCCGCCCCC   61
           || |  || |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||
Db   1783  GCCCCCACTGTGGCCGAGGCCCACGTTCACCGTGTTCCCCGTGTTTCCCTTGCACGGTGGTCCCCGGCCA  121

Qy     62  TTCCCCGTGCTGGGCCTGGACCTCCCGGGGGGGGGGGGGGGGGGGGGGGGCCTCATGAACTCCTTCCCGCA  121
           ||||||| ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   1843  TTCCCCGTGCTGGGCCTGGACTCCCGGGGGGGGGGGGGTGGGGCCGGGCCTCATGAACTCCTTCCCGCA  1902

Qy    122  CCTCAGGGTCACGCCCAGAAACCCCTGCAGGTCGGGGGTGAGCTCCAGTCCCCGGTTCTTT  181
           |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   1903  CCTCAGGGTCACGCCCAGAACCCCCTGCAGGTCGGGGGTGAGCTCCAGTCCCCGGTTCTTT  1962

Qy    182  GCCTCCCAGGGGCTGCGCGCCCAGAGTCCATTC----------------------------  211
           |||||||||||||||| ||||||||||||||||
Db   1963  GCCTCCCAGGGGCTGCGCGCCCAGAGTCCATTCCAGGTGCAGTGAGTAGGGGCCCGGGGCCA  2022

Qy    212  -----------------------------------------CAGGCCGCGCGGGGGGCGCGCC  228
                                                     |||||||||||||||||||||
Db   2023  GGCGCGCGACCAGGCGACCGCGGGCCCCACACGGGCGCCGCTTGTCTCCGCGCGGCGCGCGCC  2082

Qy    229  CCCGCCCACGCCCCCAGGCCCCGGCCCCGGAGCCCCTCCAGGTGGACTTGCTTCCCGGGCT  288
           || |||||| |||| ||||||||||| ||||||||||| || |||||||||||||||||| |
Db   2083  CCCACCCACGCCCCCAGGCCCCGGGGGGGGGGGGCCCTCCAAGTGGACTTGCTGCTTCCCGGGCT  2142

Qy    289  CGCCGGCGCGGGAGTCCGCGCCGGGCCGGCTTCTCGCCGCCCGCGCCGCCTCGCCGCCCCGT  348
           |||| || ||  ||| |||||||||||||| || ||||| |||| || || ||  || |
Db   2143  CGCCGCCCGCGCAGGAATCAGCCGCCCGGGGGGGCCGGGCCTGCTGCGGCTGCTCCCAGT  2202

Qy    349  CGCTGCGCGCGGCGCCCCCGCGCGCCCGCCGGCCCGCCCGGGGACACGCGGCCCTGAAGCA  408
           |||||| |||||||||||  || || |||||| || ||||||| ||||| |||||||||
Db   2203  GGTTACTGCACCCCCCCGCGCGCCCCGGGGGGGCCTTCACAGTGGACACAGCGCCTGAAGCA  2262

Qy    409  GCCTCCGCGGCCACCCCCTGCGCGGTGTCGGGCCCGGGCCCGGGCGGGCCCCGGC  468
           ||||| |||||||||| ||||| ||||||||||||||||||||||||||||||||
Db   2263  GCCTCCGGCGGCCACCCCCGTCTGCGCACCAGCCGCCCGAAGCGGCCGGC  2322

Qy    469  CCCGGCCTTCCGCGCCCACTATCGCGCGGGGGG  500
           |||||||||||||||||||| |||||||||||||
Db   2323  CCCGGCCGCCGCGGCCCACTATCGCGCACGAGCAGCGG  2354
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 175599)
AUTHORS   DOE Joint Genome Institute.
TITLE   Sequencing of Human Chromosome 16
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 175599)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Apr 25, 2001 this sequence version replaced gi:7689951.
            -----Genome Center
            Center: Joint Genome Institute
            Center code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 600095
            Center clone name: RPCI-11_50412

            Summary Statistics
            Consensus quality: 172472 bases at least Q40
            Consensus quality: 174610 bases at least Q30
            Consensus quality: 175032 bases at least Q20
            Estimated insert size: 190110; agarose-fp estimation
            Estimated insert size: 175199; sum-of-contigs estimation
            Quality coverage: 10.66 in Q20 bases; agarose-fp estimation
            Quality coverage: 11.57 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *   1   22409: contig of 22409 bp in length
            *   22410   2250: gap of unknown length
            *   2250   3195: contig of 9445 bp in length
            *   32510   3054: gap of unknown length
            *   31955   32055: contig of 8243 bp in length
            *   32055   116297: gap of unknown length
            * 116298   116397: contig of unknown length
            * 116398   145562: gap of 29165 bp in length
            * 145563   145562: contig of unknown length
            * 145563   175599: contig of 29937 bp in length

FEATURES   Location/Qualifiers
     source   1..175599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-50412"
                     /clone_lib="RPCI human BAC library 11"

BASE COUNT   43184 a 44157 c 44428 g 43430 t 400 others
ORIGIN

Query Match   27.8%; Score 138.4; DB 2; Length 175599;
Best Local Similarity 50.0%; Pred. No. 2.4e-07;
Matches 196; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

Qy   109 CTCCTTCCCGGCCACCTCAGGGTCACGCCCAGAACCCCTGCAGGTCGGGGGCTGAGCTCCA 168
           || || | ||||||||||||||||||| ||| | |||| ||| || || ||
Db 116137 CGCCCGGCGGCTCCGGCGGCGTAGTGTGTCGTCCGCTTCCACCCTCACCTCCCCCCTCCT 116196

Qy   169 GTCCCGGCTTCTTTTGCCTCCCCAGGGGTTGCGCCCCGGAGAGTCCCATTCCAGGCCGGCGCGCGCC 228
           || || |||| | ||||| |||||||||||||||||||| ||||| || ||
Db 116197 GCCCGGCCCGGCCCCGGCTCCTCCGGGGTCCCCCCCCCACCCCGCCGCCGCCTCCTCCTCC 116256

Qy   229 CCCGGCCCACGCCCCGGCGGCCCGGCGGGGCAGCCCCTGGACTTGCTCCCGGTGCT 288
           | |||| |||| | | || | |||||
Db 116257 CGCCCAGGGTGAGGCCGGCGCCACCTCCCTCCCTCCTCCGGCNNNNNNNNNNNNNNNN 116316

Qy   289 CGCCGCGCCCAGGAGTCCGCGCCGCGCGGGTCTGCGGCGGCCGCCGCCGCCGCCGCGCCGCCGT 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116317 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 116376

Qy   349 CGCTGCGCGCGCGCGCGCCGCGCGCGCGCGCCGCCGCTTCACGCGCGCGTGGACACCAGGAGCA 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116377 NNNNNNNNNNNNNNNNNNTCCGCGCGCCGCCGCCTCTACGGTGGACACAGGGGCCCCTGAAGCA 116436

Qy   409 GCCTCCGGCGCCCGCCCTCCGGCCACCCCGCCCAGTGTCGCGCCCCGGAGCCCGCCGGCC 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116437 GCCTCCGGCGCCCCCCTCCGGCCACCCCAGTGTCGCGCCCCGGCGCCGGCCGGAGCCGGGGCCGGGCGGCGG 116496

Qy   469 CCCGGCCTCCGCGCGCGCCCACTATCCCGCGCGGCGG 500
           |||||||||||||||||||||||||||||| |||
Db 116497 CCCGGCCTCCGCGGCGCCACTATCCCGCGCTCGG 116528

RESULT 12

AC093933

LOCUS   AC093933   86621 bp   DNA   linear   HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-5E19, *** SEQUENCING IN PROGRESS ***,
            52 unordered pieces.
ACCESSION   AC093933
VERSION   AC093933.2 GI:17941468
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 86621)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blakenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kuteshi,A., Landry,H., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Role,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 86621)

AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (12-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Dec 20, 2001 this sequence version replaced gi:15559117.
            -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: GBEK
              Center clone name: CH230-5E19
            -------------- Summary Statistics
              Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Estimated insert size: 34130; sum-of-contigs estimation
              Quality coverage: 66189 bases at least Q20
              Consensus quality: 60967 bases at least Q30
              Consensus quality: 51851 bases at least Q40
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE:This is a 'working draft' sequence. It currently
            * consists of 52 contigs.The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1005: contig of 1005 bp in length
            *   1006   1105: gap of unknown length
            *   1106   3795: contig of 2690 bp in length
            *   3796   3895: gap of unknown length
            *   3896   6097: contig of 2202 bp in length
            *   6098   6197: gap of unknown length
            *   6198   8292: contig of 2095 bp in length
            *   8293   8392: gap of unknown length
            *   8393   10938: contig of 2546 bp in length
            *   10939   11038: gap of unknown length
            *   11039   14624: contig of 3586 bp in length
            *   14625   14724: gap of unknown length
            *   14725   16339: contig of 1615 bp in length
            *   16340   16439: gap of unknown length
            *   16440   17745: contig of 1306 bp in length
            *   17746   17845: gap of unknown length
            *   17846   20378: contig of 2533 bp in length
            *   20379   20478: gap of unknown length
            *   20479   22174: contig of 1696 bp in length
            *   22175   22274: gap of unknown length
            *   22275   24039: contig of 1765 bp in length
            *   24040   24139: gap of unknown length
            *   24140   25974: contig of 1835 bp in length
            *   25975   26074: gap of unknown length
            *   26075   27147: contig of 1073 bp in length
            *   27148   27247: gap of unknown length
            *   27248   29160: contig of 1913 bp in length
            *   29161   29260: gap of unknown length
            *   29261   30732: contig of 1472 bp in length
            *   30733   30832: gap of unknown length
            *   30833   32094: contig of 1262 bp in length
            *   32095   32194: gap of unknown length
            *   32195   33617: contig of 1423 bp in length
            *   33618   33717: gap of unknown length
            *   33718   34862: contig of 1145 bp in length
            *   34863   34962: gap of unknown length
            *   34963   36546: contig of 1584 bp in length
            *   36547   36646: gap of unknown length
            *   36647   38505: contig of 1859 bp in length
            *   38506   38605: gap of unknown length
            *   38606   39758: contig of 1053 bp in length
            *   39759   39858: gap of unknown length
            *   39859   42005: contig of 2247 bp in length
            *   42006   42105: gap of unknown length
            *   42106   43452: contig of 1348 bp in length
            *   43453   43553: gap of unknown length
            *   43554   44846: contig of 1293 bp in length
            *   44847   44946: gap of unknown length
            *   44947   46475: contig of 1529 bp in length
            *   46476   46575: gap of unknown length
            *   46576   48191: contig of 1616 bp in length
            *   48192   48291: gap of unknown length
            *   48292   49384: contig of 1093 bp in length
            *   49385   49484: gap of unknown length
            *   49485   50981: contig of 1497 bp in length
            *   50982   51081: gap of unknown length
            *   51082   52733: contig of 1652 bp in length
            *   52734   52833: gap of unknown length
            *   52834   54378: contig of 1545 bp in length
            *   54379   54478: gap of unknown length
            *   54479   55702: contig of 1224 bp in length
            *   55703   55802: gap of unknown length
            *   55803   57410: contig of 1608 bp in length
            *   57411   57510: gap of unknown length
            *   57511   59250: contig of 1740 bp in length
            *   59251   59350: gap of unknown length
            *   59351   60433: contig of 1083 bp in length
            *   60434   60533: gap of unknown length
            *   60534   62025: contig of 1492 bp in length
            *   62026   62125: gap of unknown length
            *   62126   63246: contig of 1121 bp in length
            *   63247   63346: gap of unknown length
            *   63347   64824: contig of 1478 bp in length
            *   64825   64924: gap of unknown length
            *   64925   65927: contig of 1003 bp in length
            *   65928   66027: gap of unknown length
            *   66028   67879: contig of 1852 bp in length
            *   67880   67979: gap of unknown length
            *   67980   69428: contig of 1449 bp in length
            *   69429   69528: gap of unknown length
            *   69529   70934: contig of 1406 bp in length
            *   70935   71034: gap of unknown length
            *   71035   72240: contig of 1206 bp in length
            *   72241   72340: gap of unknown length
            *   72341   73906: contig of 1566 bp in length
            *   73907   74006: gap of unknown length
            *   74007   75109: contig of 1103 bp in length
            *   75110   75209: gap of unknown length
            *   75210   76292: contig of 1083 bp in length
            *   76293   76392: gap of unknown length
            *   76393   77727: contig of 1335 bp in length
            *   77728   77827: gap of unknown length
            *   77828   79159: contig of 1332 bp in length
            *   79160   79259: gap of unknown length
            *   79260   80267: contig of 1008 bp in length
            *   80268   80367: gap of unknown length
            *   80368   81774: contig of 1407 bp in length
            *   81775   81874: gap of unknown length
            *   81875   83242: contig of 1368 bp in length
            *   83243   83342: gap of unknown length
            *   83343   84618: contig of 1276 bp in length
            *   84619   84718: gap of unknown length
            *   84719   86621: contig of 1903 bp in length.
FEATURES   Location/Qualifiers
     source   1..86621
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-5E19"
BASE COUNT   19623 a   21209 c   20588 g   19594 t   5607 others
ORIGIN
        Query Match   19.5%;   Score 97.6;   DB 2;   Length 86621;

```
Best Local Similarity 50.6%;  Conservative 0;  Pred. No. 0.0072;  Indels 0;  Gaps 0;
Matches 235;                                  Mismatches 229;

Qy     37  TCCTTGCACGGTGCTGGCCCCCCCCGTCGTGCTGGCCCTGGACTCCCGGGGGTGGG          96
           |||   ||||||||||||||     |||||        |||  |||||
Db  36750  TCCCCCCCCCCCCCCCCCCCCCCCCCCGGACCAGCCGGTCCGGCCGGCCCGGC           36809

Qy     97  CGGGCCTCATGAACTCCTTCCCCGGGTCACGGCCCCAGAACCCCTGCAGGTCGG            156
           | ||   ||        |||||||  ||||        ||||||||||||||
Db  36810  CCCCCCCCACCCCCACCCCCCGCCTTCCCGGCCTTCCCCCCCCCCCCCCC             36869

Qy    157  GGGCTGAGCTCAGTCCCGGCTTCTTTGCCTCCCAGGGCTGCGCCCAGAGTCCATTCCAGGC   216
           |  ||||||||||||||||||  |||||||||||||
Db  36870  CCCACCGCGCCCACCCACCACGGGCCCCCCCCCCCCCCCCCCCCCCCCCCAC           36929

Qy    217  CGGGCCGGCCGCCCCCACGCCCAGGCCCGGCGAGCCCTTCCAGGTGGACTT             276
           |||||     |||  ||||        ||||||||||||||
Db  36930  CCCACCCCCCTCCCCCGCCGCCCCCCCCCCCCCCCCCCCCCCCCCACCC             36989

Qy    277  GCTCCGGGTCGTCGCCGCCCCAGGAGTCCGGCGCCGGCTGCTCGCGCGCGCCGCGC        336
           |    |||       |||||        |||||||||||||
Db  36990  CCCCCCGCCGTCCCCCAGTCCCCCCCCCCCCCCCCCCCCCCCCGGG               37049

Qy    337  TGCTGCCGCCGCGCTGCGCTCGGCCGCCTTCCCCGCGCGCCGCTTCACGGCTGGACACAGC   396
           |||     |||||||||  |||||||||||||
Db  37050  GCGAGGAGCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC              37109

Qy    397  GGGCCCTGAAGCAGCTTCCGCGCACCCCCGCCAGTGTGTCGGCGCCGGGCCGGGC         456
           |||          |||||  |||||||||||||
Db  37110  CCCCCTCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC                 37169

Qy    457  CGAGGCCGCGCGCGCGCCTCCGCCCACTATCGCCCGGCGGCCGG                    500
           |   ||||||||||||||||  |||||||||||||
Db  37170  CCCCCCCCCCCCCCCCCCCCCCCTACCCACCCCGCCGCGGGGCGG                  37213


RESULT 13

LOCUS       AC021596/c         93821 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-165K4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC021596
VERSION     AC021596.2  GI:9152441
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 93821)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome, clone RP11-165K4
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 93821)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
JOURNAL     Direct Submission
            Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6705503.
```

```
*      1    893:  contig of 893 bp in length
*    894    993:  gap of 100 bp
*    994   1890:  contig of 897 bp in length
*   1891   1990:  gap of 100 bp
*   1991   2888:  contig of 898 bp in length
*   2889   2988:  gap of 100 bp
*   2989   3896:  contig of 908 bp in length
*   3897   3996:  gap of 100 bp
*   3997   4900:  contig of 904 bp in length
*   4901   5000:  gap of 100 bp
*   5001   5889:  contig of 889 bp in length
*   5890   5989:  gap of 100 bp
*   5990   6921:  contig of 932 bp in length
*   6922   7021:  gap of 100 bp
*   7022   7913:  contig of 892 bp in length
*   7914   8013:  gap of 100 bp
*   8014   8920:  contig of 907 bp in length
*   8921   9020:  gap of 100 bp
*   9021   9922:  contig of 902 bp in length
*   9923  10022:  gap of 100 bp
*  10023  10968:  contig of 946 bp in length
*  10969  11068:  gap of 100 bp
*  11069  11978:  contig of 910 bp in length
*  11979  12078:  gap of 100 bp
*  12079  13022:  contig of 944 bp in length
*  13023  13122:  gap of 100 bp
*  13123  14061:  contig of 939 bp in length
*  14062  14161:  gap of 100 bp
*  14162  15097:  contig of 936 bp in length
*  15098  15197:  gap of 100 bp
*  15198  16178:  contig of 981 bp in length
*  16179  16278:  gap of 100 bp
*  16279  17172:  contig of 894 bp in length
*  17173  17272:  gap of 100 bp
*  17273  18186:  contig of 914 bp in length
*  18187  18286:  gap of 100 bp
*  18287  19182:  contig of 896 bp in length
*  19183  19282:  gap of 100 bp
*  19283  20020:  contig of 738 bp in length
*  20021  20120:  gap of 100 bp
*  20121  21043:  contig of 923 bp in length
*  21044  21143:  gap of 100 bp
*  21144  22082:  contig of 939 bp in length
*  22083  22182:  gap of 100 bp
*  22183  23111:  contig of 929 bp in length
*  23112  23211:  gap of 100 bp
*  23212  24098:  contig of 887 bp in length
*  24099  24198:  gap of 100 bp
*  24199  25112:  contig of 914 bp in length
*  25113  25212:  gap of 100 bp
```

```
* 25213  26146: contig of 934 bp in length
* 26147  26246: gap of 100 bp
* 26247  27157: contig of 911 bp in length
* 27158  27257: gap of 100 bp
* 27258  28150: contig of 893 bp in length
* 28151  28250: gap of 100 bp
* 28251  29132: contig of 882 bp in length
* 29133  29232: gap of 100 bp
* 29233  30132: contig of 900 bp in length
* 30133  30232: gap of 100 bp
* 30233  31118: contig of 886 bp in length
* 31119  31218: gap of 100 bp
* 31219  32142: contig of 924 bp in length
* 32143  32242: gap of 100 bp
* 32243  33140: contig of 898 bp in length
* 33141  33240: gap of 100 bp
* 33241  34132: contig of 892 bp in length
* 34133  34232: gap of 100 bp
* 34233  35107: contig of 875 bp in length
* 35108  35207: gap of 100 bp
* 35208  36160: contig of 953 bp in length
* 36161  36260: gap of 100 bp
* 36261  37163: contig of 903 bp in length
* 37164  37263: gap of 100 bp
* 37264  38185: contig of 922 bp in length
* 38186  38285: gap of 100 bp
* 38286  39164: contig of 879 bp in length
* 39165  39264: gap of 100 bp
* 39265  40177: contig of 913 bp in length
* 40178  40277: gap of 100 bp
* 40278  41175: contig of 898 bp in length
* 41176  41275: gap of 100 bp
* 41276  42173: contig of 898 bp in length
* 42174  42273: gap of 100 bp
* 42274  43192: contig of 919 bp in length
* 43193  43292: gap of 100 bp
* 43293  44209: contig of 917 bp in length
* 44210  44309: gap of 100 bp
* 44310  45215: contig of 906 bp in length
* 45216  45315: gap of 100 bp
* 45316  46245: contig of 930 bp in length
* 46246  46345: gap of 100 bp
* 46346  47231: contig of 886 bp in length
* 47232  47331: gap of 100 bp
* 47332  48286: contig of 955 bp in length
* 48287  48386: gap of 100 bp
* 48387  49311: contig of 925 bp in length
* 49312  49411: gap of 100 bp
* 49412  50320: contig of 909 bp in length
* 50321  50420: gap of 100 bp
* 50421  51285: contig of 865 bp in length
* 51286  51385: gap of 100 bp
* 51386  52310: contig of 925 bp in length
* 52311  52410: gap of 100 bp
* 52411  53349: contig of 939 bp in length
* 53350  53449: gap of 100 bp
* 53450  54359: contig of 910 bp in length
* 54360  54459: gap of 100 bp
* 54460  55352: contig of 893 bp in length
* 55353  55452: gap of 100 bp
* 55453  56392: contig of 940 bp in length
* 56393  56492: gap of 100 bp
* 56493  57408: contig of 916 bp in length
* 57409  57508: gap of 100 bp
* 57509  58417: contig of 909 bp in length
* 58418  58517: gap of 100 bp
* 58518  59424: contig of 907 bp in length
* 59425  59524: gap of 100 bp
* 59525  60462: contig of 938 bp in length
* 60463  60562: gap of 100 bp
* 60563  61453: contig of 891 bp in length
* 61454  61553: gap of 100 bp
* 61554  62519: contig of 966 bp in length
```

```
* 62520  62619: gap of 100 bp
* 62620  63493: contig of 874 bp in length
* 63494  63593: gap of 100 bp
* 63594  64511: contig of 918 bp in length
* 64512  64611: gap of 100 bp
* 64612  65518: contig of 907 bp in length
* 65519  65618: gap of 100 bp
* 65619  66511: contig of 893 bp in length
* 66512  66611: gap of 100 bp
* 66612  67538: contig of 927 bp in length
* 67539  67638: gap of 100 bp
* 67639  68550: contig of 912 bp in length
* 68551  68650: gap of 100 bp
* 68651  69551: contig of 901 bp in length
* 69552  69651: gap of 100 bp
* 69652  70552: contig of 901 bp in length
* 70553  70652: gap of 100 bp
* 70653  71591: contig of 939 bp in length
* 71592  71691: gap of 100 bp
```

```
Query Match   19.4%;  Score 96.8;  DB 2;  Length 93821;
Best Local Similarity   51.6%;  Pred. No. 0.0085;
Matches  235;  Conservative  0;  Mismatches 219;  Indels  1;  Gaps  1;

Qy    33 TGTTTCCTTGACGCTGCTGGCCCCCCCCCCTTCCCGTGCTGGGGCCTGGACTCCGGGGGGG 92
          ||   |  || ||||  |  |||  | ||||||||| |||   |||| ||||||| ||| |
Db 90811 TTTTTTTTTTTTTTTTTTACACGGCCACGCCCCCCCCCCCCCCCACGCCCCACCCCCCCGC 90752

Qy    93 TGGGCGGCCCTCATTGAACTCCTTCCCGCGCCACCTCAGGGTGCACGCGCCCAGAACCCCC 152
          | |  ||||  |   |  |  |   |||  || |  | |||  | ||||| |   |||| ||
Db 90751 CCCCCCCCCCCCCCGCCCCCCCCCCCCNGCACGCCCCCCNGCACGCGCCCCCCCCCCCGC 90692

Qy   153 TCGGGGCTGAGCTCCAGTCCCCGCTTCTTTTGCCTC-CCAGGAGTCGCGCGCCCCAGAGTCCCATTC 211
          |||   |     |||  ||  ||  |   |  | ||| |  | || |||||  |  |  |  ||||
Db 90691 CCGGCCCCCCGGGGNNCACCCCCGGGGNNCACCCCCGCGCCCCGGCCCCGCCCCGCG 90632

Qy   212 CAGGCCGCGCCGGCGCCCCGCGCCCCCCCACGCCCCAGGCCCCCGGGCCGAGCCCTCCAGGTG 271
          |||  ||| || | | ||||  ||  |  | |  |||| | |||||  ||| || ||  |  |
Db 90631 GCCGCCCCCCCCCCGCCCCCCCCCCCCCCCCCGCCCCCCCCGCAGGCCC 90572

Qy   272 GACTTGCTCCCGTGCTCGCGCCCCGCCAGGAGTCGCGCCGCGGGGTTGCTCGTCGGGGCTGCC 331
          ||| ||| ||| |  | ||||||  |  || | | |||| |||  | || || |||| || ||
Db 90571 GCCCGCCCCCCCCCCCCCCCCCCGGGGGGGCCCGCGCCCCCCGCCGCCACCCCGGGAC 90512

Qy   332 GCCGCTCGCTGCCGCCGTGCGCTGCGGCCCCGCCCCGGGGCCCCCCTCTACGGTGGAC 391
          ||||  |    ||||  | || ||  || |||  ||||  |||||| |||| |||| |
Db 90511 GCGCGGCGCGGCGGCCCCCCCCCCCGCGCNCGCGGGCCCCCCCCCCCGGCACCCCGGAC 90452

Qy   392 ACAGCGGCGCCTGAAGCAGGCCTCGGCGCCACCCCCGCCCCAGTGTCGCGGGCCC 451
          |       | | | || |  | |||  | ||  | | || |  | ||| || |||
Db 90451 CCCNNCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCGCCCCCCNCC 90392

Qy   452 GCGGCCGAGGCGGCGCCCCCCCCCCCCCCCCGCCTCCGCGCCAC 486
          | |   | |  ||| || ||||||||||||| || | ||||||
Db 90391 CCGCGCCGCCGGCCGCCCCCCCCCCCCCCCCGCCCCCGCGCCCCCC 90357
```

```
RESULT 14
AC084064/c
LOCUS       AC084064        197669 bp    DNA     linear   HTG 25-APR-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-467I17, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC084064
VERSION     AC084064.4  GI:13786424
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197669)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 197669)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 25, 2001 this sequence version replaced gi:1152453.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-11_467I17
            -----
            Summary Statistics
            Consensus quality: 158774 bases at least Q40
            Consensus quality: 174343 bases at least Q30
            Consensus quality: 180875 bases at least Q20
            Estimated insert size: 163300; agarose-fp estimation
            Estimated insert size: 195769; sum-of-contigs estimation
            Quality coverage: 5.93 in Q20 bases; agarose-fp estimation
            Quality coverage: 4.94 in Q20 bases; sum-of-contigs estimation.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 20 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *        1     1297: contig of 1297 bp in length
          *     1298     1397: gap of unknown length
          *     1398     2963: contig of 1566 bp in length
          *     2964     3063: gap of unknown length
          *     3064     4476: contig of 1413 bp in length
          *     4477     4576: gap of unknown length
          *     4577     5741: contig of 1165 bp in length
          *     5742     5841: gap of unknown length
          *     5842     7080: contig of 1239 bp in length
          *     7081     7180: gap of unknown length
          *     7181     8247: contig of 1067 bp in length
          *     8248     8347: gap of unknown length
          *     8348    10242: contig of 1895 bp in length
          *    10243    10342: gap of unknown length
          *    10343    12553: contig of 2211 bp in length
          *    12554    12653: gap of unknown length
          *    12654    15376: contig of 2723 bp in length
          *    15377    15476: gap of unknown length
          *    15477    20581: contig of 5105 bp in length
          *    20582    20681: gap of unknown length
          *    20682    25105: contig of 4424 bp in length
          *    25106    25205: gap of unknown length
          *    25206    29871: contig of 4666 bp in length
          *    29872    29971: gap of unknown length
          *    29972    35023: contig of 5052 bp in length
          *    35024    35123: gap of unknown length
          *    35124    41897: contig of 6774 bp in length
          *    41898    41997: gap of unknown length
          *    41998    51710: contig of 9713 bp in length
          *    51711    51810: gap of unknown length
          *    51811    62531: contig of 10721 bp in length
          *    62532    62631: gap of unknown length
          *    62632    77530: contig of 14899 bp in length
          *    77531    77630: gap of unknown length
          *    77631   112421: contig of 34791 bp in length
          *   112422   112521: gap of unknown length
          *   112522   142859: contig of 30338 bp in length
          *   142860   142959: gap of unknown length
          *   142960   197669: contig of 54710 bp in length.
FEATURES             Location/Qualifiers
     source          1..197669
                     /organism="Homo sapiens"
```

```
BASE COUNT    53353 a  42514 c  43922 g  53642 t   4238 others
ORIGIN
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-467I17"
                     /clone_lib="RPCI human BAC library 11"

Query Match    19.0%;  Score 94.8;  DB 2;  Length 197669;
Best Local Similarity  47.9%;  Pred. No. 0.011;
Matches  186;  Conservative 0;  Mismatches 202;  Indels 0;  Gaps 0;

Qy    111 CCTTCCCGCACCTCAGGGTCACGCGCCAGAGAACCCCCTGCAGGTCGGGGGGGGTCCAGT 170
          ||||  | | || ||  |  | | |  |||  | || ||  ||  |  |||||| ||
Db    854 CCCNCNGCCCGCGCCGCNCCCCCGCGNNCCCNCNGCCCCCCCCCCCCCNCCCCCCN 795

Qy    171 CGCCCACGCCCCAGGCCCGCGGGCGCGAGCCCCTCCAGGTGGACTTGCTCCGCGGTGCTCG 230
          | |  | | |  || | || |||  ||  |  | ||  ||  | ||| |  | | || ||
Db    794 NCCCCCNCCCCCNCNCCCGGNCCCCCCGGGCCCCCNCCCCCCCNNCCNGCCCCCCCCGGNCC 735

Qy    231 CGCCCACCCCAGGCCGCGGCGCGGCGGCGGCGCGCCCTCCAGGTGGACTTGCTCCGCGGTC 290
          | ||  | | ||| | | || | | | || | || || ||  ||  | ||  |  | || |
Db    734 CCCNNGNCNCCCGGCCCCCCCCGGNCCCCCNCCCCCCCNNCCCCCCCGGNCCCCCCGGNCCC 675

Qy    291 CCGCCGCCCCAGGAGAGTCCCGCGCGCGGCGGCTGCGGCGGCGGCTGCGTGCGGCGTCG 350
          | ||  |  | || | || | || | |  | | |  |  | || | | || || | | |
Db    674 CCCCGCCCCCCGGNNCCCCCCCCCNNCCCCCCCCCCCCGNGCCCCCNGCCCCCGCCC 615

Qy    351 CTGCCGCGCGCCCTGCCCCGCCGCCTTACGGTGGACACGGCGCCTGAAGCAGC 410
          | ||  | | |  |  | || |  |  || |  | || | | | || ||  | |
Db    614 CCCCGNCCCGCCCGCCCCGCNCCCGCGCCCCCCCCCCCCCGCCCNNGGNCCC 555

Qy    411 CTCCGGGCGCGGCCCTTCCGCCACCCCCCAGTGTGCGGCGCGCGGGCGACGCGCCCC 470
          |  | | | | || | ||| | || || || | | | |  | |  | |   | ||||
Db    554 CCCNCCCCCCCCNCCCCCCCCCCNNCCGNCCCGCCCCNCCNGNGCCGCCCCNNGGNCCC 495

Qy    471 CCGCCTCCGCCCGCCACTATCGCCGCCGGC 498
          ||| | || |  | || | | ||| | |
Db    494 CCCCCCCGCCCCGCCCCCGCCCCCGCCNCC 467


RESULT 15
LOCUS       AP003300              137878 bp    DNA     linear   HTG 28-NOV-2001
DEFINITION  Oryza sativa chromosome 1 clone P0700A11, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003300
VERSION     AP003300.2  GI:17132529
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa (cultivar:Nipponbare) DNA, clone:P0700A11.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 137878)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Nov 27, 2001 this sequence version replaced gi:13027330.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES             Location/Qualifiers
```

source   1. .137878
                /organism="Oryza sativa"
                /cultivar="Nipponbare"
                /db_xref="taxon:4530"
                /chromosome="1"
                /clone="P0700A11"

BASE COUNT   40322 a   29702 c   29543 g   38111 t   200 others
ORIGIN

Query Match   18.9%;   Score 94.6;   DB 2;   Length 137878;
Best Local Similarity   50.3%;   Pred. No. 0.013;
Matches 232;   Conservative   0;   Mismatches 229;   Indels   0;   Gaps   0;

Qy   38 CCTTGCACGCTGCTGGCCCCCCCCCCTTCCCGGTGCTGGGGTGGGACTCCCGGGGGGTGGGG 97
          || || | ||||| ||| || ||| | ||
Db 113898 CCCCCCCCCCGACACACCCGGCCCCACCCCCAGCCGCGCCCTTCCCCCCAAGA 113839

Qy   98 GGCCTCATGAACTCCTTCCCGCCCACCTCAGGGGTCACGCCCAGAAACCCCCTGCAGGTCGGG 157
          | | || || | || ||| || || || || |||
Db 113838 CGGGGCGCCGACCCCGCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 113779

Qy   158 GCTGAGGCTCCCAGTTCCCCGCCTTTCTTTGCCTCTCCAGGGGCTGCGGCCCAGAGTCCATTCCAGGCC 217
          | |||||| | ||||| | ||| || ||
Db 113778 CCCCCCCCGCGCCCCCCGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 113719

Qy   218 GCGCCGCGGCGCCCCCCCGCCCCACGCCCCCAGGCGCCCCGGGCGGCCCCTCCAGGTGGGACTTG 277
          | |||| | |||||||| | ||||| | | |||| | |||| ||
Db 113718 CCCCCCCCCGC|||||||CTCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCC 113659

Qy   278 CTCCCGGTGCTCGCGCGCGCCCAGGAGTCCGCCCGCGGCTGCCTGCGGCGCCGGCTGCCCGCGGCT 337
          | ||| | | |||| | | || || ||| || | ||| | | || ||
Db 113658 CCCCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCGCCCCCCTTCCC 113599

Qy   338 GCTGCCCGCCGTGGCTGCGCGCCGCCCCCGCGCCCCCTGCCCGCGCCTTACGGTGGACACAGCG 397
          |||| | || || ||||||| | | | | | ||||||| | |||| | ||
Db 113598 CCCCGCCCCCTCCCTCCCCCCCCCCCCCCCCCCCCCCTTCGCCCCCCTCCCCCCC 113539

Qy   398 GCCCTGAAGCAGCCTCCCGGCGCCCCCCTCCGGCCCACCCCAGTGTCGGGGCCGCCGCGGCC 457
          || | || || ||| ||||| ||||| || || | || | ||||
Db 113538 GCTCTCCCCTCCCCCCCTCTTTCCCCCCCCCCCCCCCCAGTGTCGGGGGGCCCGCGCGGCCC 113479

Qy   458 GAGGCCGCGCCCCCCGCCTTCCGCGCGCCCACTATCCGCGCGGC 498
          | ||||||||||||| | | |||||| | | | ||
Db 113478 GCCCCCGGGCCCCCCGCCCCCCGCCCTCCCCGGCCCCCCCCCTCCCC 113438

Search completed: October 10, 2002, 17:41:15
Job time : 990.867 secs

```
GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                 (without alignments)
                 14616.225 Million cell updates/sec

Title:           US-09-489-101A-8_COPY_1_500
Perfect score:   500
Sequence:        1 gggacgtgagccgctgcgcc.........aaatgttatctgatccagtg 500
Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched:        1797656 seqs, 1046326293 residues

Total number of hits satisfying chosen parameters:          3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htgo_inv:*
```

SUMMARIES

```
Result
No.    Score  Query  Match Length  DB    ID                 Description
-----  -----  -----  ------------  --    --------           -----------

  1    500    100.0     1860       9     AX201587   Sequence
  2    500    100.0     1860       9     X98260     H.saplens m
  3    416.8   83.4     2037      10     HSMPP11
  4    411.8   82.4     1995      10     MUSMIDA1
  5    400.2   80.0     1801      10     U53208     Mus musculu
  6    193.2   38.6    1112661    9     AF118853   Rattus no
  7    193.2   38.6     196130    2     AC004668   Homo sapi
  8    172     34.4    1189242    2     AC093701   Homo sapi
  9    167.4   33.5     134550    9     AC023284   Mus muscu
 10    167.4   33.5     164268    9     AC024402   Homo sapi
 11     93     18.6     189242    2     AC023284   Mus muscu
 12     80.2   16.0      2248     3     AY069523   Drosophil
 13     73.2   14.3     37838     8     Z97992     S.pombe chr
 14     71.4   14.3       680    11     AL593453   Meleagris
 15     71.4   14.3       680    11     AL593453   Meleagris
 16     60.4   12.1      1876     6     AX073252   Sequence 14
 17     53     10.6     19011     8     I66494     Sequence
 18     52     10.4      1802     6     AX073012   Sequence
 19     51.8   10.4      2222     8     SCYGR285C  S.pombe c
 20     51.8   10.4      3083     6     Z73070     S.cerevisia
 21     51.8   10.4      3083     6     AR075303   Sequence
 22     51.8   10.4      3083     6     I66541     Sequence 1
 23     51.8   10.4      3083     8     SCU01      S.cerevisia
 24     51.8   10.4     35400     8     SCCHV1IRA  S.cerevisia
 25     44.4    8.9   1000000     9     AP000070   Homo sapi
 26     43.6    8.7     87538     9     AL158200   Human DNA
 27     43      8.6     34867     8     AP002030   Arabidops
 28     43      8.6     77715     8     AC094734   Rattus no
 29     42.6    8.5    152528     2     AC016063   Homo sapi
 30     42.6    8.5    206454     9     AC018737   Homo sapi
 31     42.4    8.5       712     1     MG001724   Mycoplasma
 32     42.4    8.5      8160     9     U39681     Mycoplasma
 33     42.4    8.5     56952     2     AC018324   Drosophil
 34     42.4    8.5    163093     3     AC010055   Drosophil
 35     42.4    8.5    243075     2     AL591805   Mus muscu
 36     42.4    8.5    313115     3     AE003593   Drosophil
 37     42.2    8.4     77715     8     NCB24P7    Neurospor
 38     42      8.4      3234     3     AF356828   Plasmodiu
 39     41.2    8.2      2041     1     U72998     Borrelia bu
 40     41.2    8.2     30800     1     BBU72998   Borrelia
 41     41.2    8.2    109398     9     AE001579   Borrelia
 42     41      8.2     97225     2     AC044817   Homo sapi
 43     41      8.2    218228     2     AC099417   Rattus no
 44     41      8.2    224448     2     AC021476   Mus muscu
 45     40.8    8.2      2069    23     E10125     DNA encodin
```

ALIGNMENTS

```
RESULT 2
HSMPP11

LOCUS       HSMPP11     1860 bp    mRNA    linear   PRI 08-JAN-1997
DEFINITION  H.sapiens mRNA for M-phase phosphoprotein, mpp11.
ACCESSION   X98260
VERSION     X98260.1  GI:1770453
KEYWORDS    M phase phosphoprotein; MPP gene.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1860)
AUTHORS     Matsumoto-Taniura,N., Pirollet,F., Monroe,R., Gerace,L. and
            Westendorf,J.M.
TITLE       Identification of novel M phase phosphoproteins by expression
            cloning
JOURNAL     Mol. Biol. Cell 7 (9), 1455-1469 (1996)
MEDLINE     97039687
REFERENCE   2  (bases 1 to 1860)
AUTHORS     Westendorf,J.M.
TITLE       Direct Submission
JOURNAL     Submitted (03-JUN-1996) J.M. Westendorf, INSERM U366, DBMS/CS-CENG,
            17 rue des Martyrs, F- 38054 Grenoble Cedex 9, FRANCE
FEATURES             Location/Qualifiers
     source          1..1860
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="14"
                     /sex="male"
```

Query Match   100.0%;   Score 500; DB 6; Length 1860;
Best Local Similarity   100.0%;   Pred. No. 1.3e-123;
Matches 500; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 GGGACGTGAGCCGCCGCTGCGCCCACCGGGCTAGACCCGGCGCCATCATCTGCTTCTGCCAA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGACGTGAGCCGCCGCTGCGCCCACCGGGCTAGACCCGGCGCCATCATCTGCTTCTGCCAA 60

Qy         61 GCGCCGCGGACGGCCGGGGCGCCACCGGGCCATCACCACCGCTCTCGACCTCTGCCCTCTT 120
Db         61 GCGCCGCGGACGGCCGGGGCGCCACCGGGCCATCACCACCGCTCTCGACCTCTGCCCTCTT 120
```

(sequence alignment continues)

```
polyA_signal   1779..1784
polyA_site     1845
```

```
LOCUS       MUSMIDA1                2037 bp    mRNA    linear   ROD 13-FEB-1999
DEFINITION  Mouse mRNA for MIDA1, complete cds.
ACCESSION   D63784
VERSION     D63784.1  GI:1060924
KEYWORDS    .
SOURCE      Mus musculus cell-line Friend erythroleukemia cell DS19/3 cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2037)
  AUTHORS   Shoji,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1995) Wataru Shoji, Institute of Development,
            Aging and Cancer Tohoku University, Cell Biology; 4-1 Seiryomachi
            Aoba-ku, Sendai, Miyagi 980-77, Japan (Tel:022-274-1111(ex.3462),
            Fax:022-272-5081)
REFERENCE   2  (bases 1 to 2037)
  AUTHORS   Shoji,W., Inoue,T., Yamamoto,T. and Obinata,M.
  TITLE     MIDA1, a protein associated with Id, regulates cell growth
  JOURNAL   J. Biol. Chem. 270 (42), 24818-24825 (1995)
  MEDLINE   96027574
FEATURES             Location/Qualifiers
     source          1..2037
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="Friend erythroleukemia cell DS19/3"
     CDS             113..1978
                     /codon_start=1
                     /product="MIDA1"
                     /protein_id="BAA09854.1"
                     /db_xref="GI:1060925"
                     /translation="MLLLPSAAEGQGTAITHALTSASSVCQVEPVGRWFEAFVKRRNR
                     NASTSFQELEDEKKELSEESEDEELQLEEFPMLKTLDPKDWKNQDHYAVLGLGHVRYTA
                     TQRQIKAAHKAMVLKHHPDKRKAAGEPIKEGDNDYFTCITKAYEMLSDPVKRRAFNSV
                     DPTFDNSVPSKSEAKDNFFQVFSPVFERNSRWSNKKNVPKLGDMNSSFEDVDAFYSFW
                     YNFDSWREFSYLDEEEKKEAECRDERKWIEKONRATRAQRKKEEMNRIRTLVDNAYSC
                     DPRIKKFKEEEKKAKEAEARRKEQEAEKEQRQAELEAVRLAKEKEEEEVRQQ
                     ALLAKEKDIQKAIKKEROKLRNSCKSWNHFSDNEADRVKMMEEVEKLCDRLELASL
                     QGLNEILASSREVGKRAALEKQIEEVNEQMRREKEEADARMRQASKNAEKSTGSGSG
                     SKNWSEDDLQLLIKAVNLFPAGTNSREVIANYMNIHSSSGVKRTAKDVISEKAKSLQK
                     LDPHQKDINKAFDKFKKEHGVASQADSAAPSERFEGPCIDSSTPWTTEEQKLLEQAL
                     KTYPVNTPEREKIAEAVPGRIKKDCMRRYKELVEMVKAKKAAQEQVLNASRARK"
     polyA_signal    2012..2017
BASE COUNT      716 a    378 c    522 g    421 t
ORIGIN
```

```
Query Match             83.4%;  Score 416.8;  DB 10;  Length 2037;
Best Local Similarity   91.2%;  Pred. No. 3e-101;
Matches  454;  Conservative   0;  Mismatches  42;  Indels   2;  Gaps   1;
```

```
Qy      2 GGACGTGAGCCCCTGCCCCCACCGGGCTAGACCCGGCGGCCCATCAATGCTGCTTCTCCCAAG 61
          ||||| |||||||||||||||||| |||||||  |||||||||| |||||| |||||
Db     72 GGCCGTGAGCCCCGGGCCGTCCGCGGCGACCAGCCCC- -AGCCAGCATGCTGCTTCTCCGCGAG 129
Qy     62 CGCCGCGGACGGCCGCGGGGCACCCGCCATCACCCACCCCACCTCTGCCCTCTACACTCTG 121
          ||||| |||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db    130 CGCCGCGGAGGCGCCATCACCCATCACCCACGCTCTGACCCTCGCCCTCTCAGTCTG 189
Qy    122 TCAAGTTGAACCTGTGGGGAAGAATGGTTGAAGCTTTTTGTTAAGAGGAGAAACAGAAATGC 181
          ||||||||||| ||| |||||||| |||||||||||| |||||||||||||||||||| ||
Db    190 TCAAGTTGAACCTGTGGGAAGATGGTTTGAAGCTTTTTGTTAAGAGGAGAAACAGAATGC 249
Qy    182 TTCTGCCTCTTTTCAGGAACTGGAGGATAAGAAGACAGAGAGTTATCCGAGGAGTCAGAAGA 241
          |||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    250 TTCTACATCTTTTCAGGAACTGGAGAATAAGAAGAGTTGTCAGAGAGAATCAGAGAGTGA 309
Qy    242 AGAATTGCAGTTGGAAGAGTTTCCCATGCTGAAAACACACTTGATCCCAAGAACTGGAAGA 301
          |||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    310 AGAGTTACAGTTAGAAGAATTCCCCATGCTGAAAACACTTGATCCCAAAAGATTGGAGAA 369
Qy    302 CCAAGATCATTATATGCAGTTCTTGGACTTGGCCCATGTGAGATACACAGCGTCACACAGACA 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
LOCUS       MMU53208                1995 bp    mRNA    linear   ROD 10-APR-1996
DEFINITION  Mus musculus zuotin related factor (ZRF1) mRNA, complete cds.
ACCESSION   U53208
VERSION     U53208.1  GI:1256831
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1995)
  AUTHORS   Hughes,R., Chan,F.Y., White,R.A. and Zon,L.I.
  TITLE     Cloning and chromosomal localization of a mouse cDNA with homology
            to the Saccharomyces cerevisiae gene zuotin
  JOURNAL   Genomics 29 (2), 546-550 (1995)
  MEDLINE   96115610
REFERENCE   2  (bases 1 to 1995)
  AUTHORS   Zon,L., Hughes,R., Chan,F.Y. and White,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-1996) Leonard Zon, HHMI, Children's Hospital, 300
            Longwood Ave, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..1995
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
     gene            69..1613
                     /gene="ZRF1"
                     /note="ZRF1"
     CDS             69..1613
                     /gene="ZRF1"
                     /note="similar to DnaJ encoded by GenBank Accession Number
                     L16953"
                     /codon_start=1
                     /product="zuotin related factor"
                     /protein_id="AAC52486.1"
                     /db_xref="GI:1256832"
                     /translation="MLLLPSAAEGQGTAITHALTSASSVCQVEPVGRWFEAFVKRRNR
                     NASTSFQELEDKKELSEESEDEELQLEEFPMLKTLDPKDWKNQDHYAVLGLGHVRYTA
                     TQRQIKAAHKAMVLKHHPDKRKAAGEPIKEGDNDYFTCITKAYEMLSDPVKRRAFNSV
                     DPTFDNSVPSKSEAKDNFFQVFSPVFERNSRWSNKKNVPKLGDMNSSFEDVDAFYSFW
                     YNFDSWREFSYLDEEEKKEAECRDERKWIEKONRATRAQRKKEEMNRIRTLVDNAYSC
                     DPRIKKFKEEEKKAKEAEARRKEQEAEKEQRQAELEAVRLAKEKEEEEVRQQ
                     ALLAKEKDIQKAIKKEROKLRNSCKSWNHFSDNEADRVKMMEEVEKLCDRLELASL
                     QGLNEILASSTREVGKRAALEKQIEEVNEQMRREKEEADARMRQASKNAEKSTGSGSG
                     SKNWSEDDLQLLIKAVNLFPAGRNSRWDVIANYMNIHSSSGVKRTAKDVISEVRGPKS
                     WGSSKR"
BASE COUNT      701 a    364 c    516 g    414 t
ORIGIN
```

```
Query Match             82.4%;  Score 411.8;  DB 10;  Length 1995;
Best Local Similarity   90.3%;  Pred. No. 6.6e-100;
Matches  440;  Conservative   0;  Mismatches  47;  Indels   0;  Gaps   0;
```

```
Db    370 TCAAGATCATTATGCAGTTCTTGGACTTGGCCCAGTGCAGTACACAGGCTACACAGACA 429
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Qy    362 GATCAAAGCAGCTGCTATAAAGCAATGGTTTTAAAACATCACCCAGAGCAGCAAAAGCAGC 421
          |||||||||||||||||||||||||||||||||| ||||| ||  |||||||||| ||
Db    430 GATCAAAGCAGCTCATAAAGCAATGGTCTTGAAAACATCATTCCAGACAAACGGAAAGCGGC 489
Qy    422 TGGTGAACCAATAAAAGAAGGAGAGATAATGACTAATGACTACTTCACTTGCATAACTAAAGCTTATGA 481
          ||| |||| ||||||||||||||||| |||||| |||||||| |||||||||||| ||||
Db    490 TGGGGAACCCATCAAAAGAAGGAGAGACAATGACTACTTCACTTGCATAACTAAACTAAAGCTTATGA 549
Qy    482 AATGTTATCTGATCCAGT 499
          |||||| |||||||||||
Db    550 AATGTTGTCTGATCCAGT 567
```

```
Qy     13 GCTGCGGCGCCCACCGGCGTAGACCCGGCGCCCATCAATGCTGCTTCTGCCAAGCGCGCCGGACG 72
          ||  ||||| |  ||||| |||||||||||||||| ||||||||||||||||||||||||||
Db     37 GCGCGCGGCGTCGGCGACCAGCCCCCAGCCAGCATGCTGCTTCTGCCAGCGCGCCGGGAGG 96
```

RESULT 5

```
LOCUS       AF118853                1801 bp    mRNA    linear   ROD 26-APR-2001
DEFINITION  Rattus norvegicus gliosarcoma-related antigen MIDA1 (MIDA1) mRNA,
            partial cds.
ACCESSION   AF118853
VERSION     AF118853.1  GI:5577976
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1801)
  AUTHORS   Okada,H., Attanucci,J., Giezeman-Smits,K.M., Brissette-Storkus,C.,
            Fellows,W.K., Pollack,I.F., Pogue-Geile,K., Lotze,M.T., Bozik,M.E.
            and Chambers,W.H.
  TITLE     Immunization with an antigen identified by cytokine tumor
            vaccine-assisted SEREX (CAS) suppressed growth of the rat 9L glioma
            in vivo
  JOURNAL   Cancer Res. 61 (6), 2625-2631 (2001)
   MEDLINE   21182273
   PUBMED    11289140
REFERENCE   2  (bases 1 to 1801)
  AUTHORS   Okada,H., Attanucci,J., Giezeman-Smits,K.M., Brissette-Storkus,C.,
            Fellows,W.K., Pollack,I.F., Pogue-Geile,K., Lotze,M.T., Bozik,M.E.
            and Chambers,W.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1999) Division of Neuro-Oncology, Neurosurgery,
            University of Pittsburgh Medical Center, Pittsburgh Cancer
            Institute, Suite 802 3471 Fifth Ave., Pittsburgh, PA 15213, USA
FEATURES             Location/Qualifiers
     source          1..1801
                     /organism="Rattus norvegicus"
                     /strain="Fischer 344"
                     /db_xref="taxon:10116"
     gene            1..>1801
                     /gene="MIDA1"
     CDS             114..>1801
                     /gene="MIDA1"
                     /codon_start=1
                     /product="gliosarcoma-related antigen MIDA1"
                     /protein_id="AAD45407.1"
                     /db_xref="GI:5577977"
                     /translation="MLLLPSAAEGQGYTAITHALTSASAVCQVEPVGRWFEAFVKRNR
                     NASTSFQELEDKKELSEESEDELQLEEFPMLKTLDPKDWKNQDHYAVLGLGHVRYKA
                     TQRQIKAAHKIMVLKIHPDKRAAGEPIKEGDNDYFYCITKAYEMLSDPVKRRAFNSV
                     DPTFDMNSVPSKSEAKENFFQVFSPSVFERNSRRWSNKNWNPKLGDMNSSPEDVDAFYSFW
                     YNFDSWREFSYLDEEREKEAECRDERRWIEKONRATRAQRKKEMNRIRTLVDNAYSC
                     DPRIKFKFEEGKAKKEAEARRKEQEARKEKOQRQAELEAVRLAKEKEEEEVRQQ
                     ALLAKEKEIOKKAIKKERQKLRNSCKNWNHFSDNEADRVRKMMEEVERKLCDRLELASL
                     QCLNEILASSTREVGRAALEKOIEEVNELMRKEEEADARMRQASKNAEKSTGGSGSG
                     SKNWEBDDLLQLLIKAVNLFPAGTNSRWEVIANYMNIHSSSGVKRTAKDVIGKAKSLOK
                     LDPHQKDDINKKAFDKFKKEHGVAPOQDSAAPSERFEGPCIDSIPWTTEEQKLL"
BASE COUNT      635 a    342 c    462 g    362 t
ORIGIN
```

```
Qy   73 GCCGCGGGCACCGCCATCACCCACGCGCTGCCTTCCGCCTTCTGCCTCTGCACTCTTGTCAAGTTGAAC 132
        ||| |||||||||||||||||| || ||||||||||||||||||||||||| |||| ||||||||||||||
Db   97 GCCAGGGCACCGCCATCACCCACGCGCTCGCCTCCGCCTTCTCAGTTGTCAAGTTGAAC 132

Qy  133 CTGTGGGAACGAATGGTTGTGAACCTTTGTTAAGAGGAGGAGGAAAAATGCTTCTGCCCTCTT 192
        || ||||||||| |||||||||||||||||||||||||||||||| ||||||||||||| || |
Db  157 CTGTGGGAAGATGGTTGTGAACCTTTGTTAAGAGGAGGAGGAAAAAGTCTTCTACATCTT 216

Qy  193 TTCAGGAACTGGAGGATAAGAAAGAGTTATCCGAGAGATCAGAGAGATTGCAGT 252
        ||||||||||||||||| ||||||| |||| ||||||| ||||||| ||||||||
Db  217 TTCAGGAACTGGAAGATAAGAAAGAGTTGTCGAGAGATCAGAGAGATCAGAGAGTTACAGT 276

Qy  253 TGGAAGAGTTTCCCATGCTGAAAAAACACTTGATCCCAAAGACTGGAAGAACCAAGATCATT 312
        || ||| || |||||||||| ||||||||||||||||||||| |||||||||||| |||||
Db  277 TAGAGAATTCCCCATGCTGAAAAACACTTGATCCCAAAGATTGGAAGAATCAAGATCATT 336

Qy  313 ATGCAGTTCTTGGACTTGGCCATTGAGATACACAGAGCTACACAGAGACACAGAGATCAAAGGCAG 372
        |||||||||||||||||||||||||| || ||||||||||| || ||||||||||||||||||||
Db  337 ATGCAGTTCTTGGACTTGGCCACGTCAGATACACACTTACACAGAGACACAGAGATCAAACCAG 396

Qy  373 CTCATAAAAGCAAATGGGTTGTTAAAACATCACCCAGACAAAACGGAAAGCAGCTGGTGAACCAA 432
        || || |||||||||||||||| ||||||||||||| ||||||||||||| ||||||||||||||
Db  397 CTCATAAAAGCAAATGGGTTGTTTGAAAACATCACCCAGACAAAACGGAAAGCAGCTGGTGAACCCA 456

Qy  433 TAAAAGAAGGAGATAATGACTACTTCACTTGCATAAAGCTTATGAAAGTTTACTG 492
        |||||||||||||||||| ||||||||||||||||||| |||||||| ||| |||
Db  457 TCAAAGAAGGAGACAATGACTACTTCACTTGCATAACTAAAGCTTATGAAAATGTTGTCTG 516

Qy  493 ATCCAGT 499
        |||||||
Db  517 ATCCAGT 523
```

RESULT 6

```
LOCUS       AC004668              112661 bp    DNA     linear   PRI 03-FEB-2000
DEFINITION  Homo sapiens BAC clone CTA-27603 from 7q22-q31.1, complete
            sequence.
ACCESSION   AC004668
VERSION     AC004668.1  GI:3115345
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112661)
  AUTHORS   Rohlfing,T., Wohldmann,P., Antoniou,B., Bauer,C. and O'Neal,D.
  TITLE     The sequence of Homo sapiens BAC clone CTA-27603
```

```
Query Match             80.0%;  Score 400.2;  DB 10;  Length 1801;
Best Local Similarity   89.0%;
Matches  444;  Conservative    0;  Mismatches   53;  Indels    2;  Gaps    1;

Qy    2 GGACGTTGAGCCGCGCGCTCGCCCACCGGCGGCTGAGCCGCGGCCGCCGCCATCATGCTGCTTCTGCCCAAG 61
        ||| ||||||||||||||||||| |||||||||||||||||||||||||||| |||||||||| ||||||
Db   73 GGACGAGCAGCCGCGCGGCGCTCCGCGGAACAGCGTCC--AGCCAGCATGCTGCTTCTGCCCAG 130

Qy   62 CGCCGCGGGACGGCCGCGGGGCCCACCGCCATCACCCACCGACGTCGACCTCTGCCTCTACACTCTG 121
        ||||||||| |||||||||||||||||| ||||||||| ||| ||||| |||||||| ||||||||
Db  131 CGCCGCGGAGCCCAGGGCACCGCCATCACCCACCGCTCTGACCTCCGCCTCTGCCAGTCTG 190

Qy  122 TCAAGTTGAACCTGTGGGAACAATGGGTTTGTGAACGTTTGTTAAGAGGAGGAGGAAAAATGC 181
        |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||||
Db  191 TCAAGTCGAACCGTGGGAACAATGGTTTGTGAACGTTTCGTTAAGAGGAGGAGAAAAATGC 250

Qy  182 TTCTGCCCTCTTTCTTCAGGAACTGGAGGATAAGAAAGAGTTATCCGAGGAATCAGAGAGTGA 241
        |||| ||||||| |||||||||||||||||||||||||||||||| |||||||||||||||
Db  251 TTCCACATCTTTTCAGGAACTGGAAGATAAGAAAGAGTTGTCGAGAGATCAGAGAAGATGA 310

Qy  242 AGAATTGCAGTTGGAAGAGTTTCCCATGCTGAAAACACTCAGACACAGAGACTGGAAGAA 301
        ||||||||||| ||||||||| |||||||||| ||||||||||||||||||||||||||
Db  311 AGAGTTACAGTTAGAAGAGTTTCCCATGCTGAAAAACACTTGATCCCAAAGACTTGGAAGA 370

Qy  302 CCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGATCAAGGCTACACAGAGACA 361
        ||||| |||||| |||||||||||||||||||||| |||||| ||| ||||||||||||
Db  371 TCAAGATCATTATGCAGTTCTTGGACTTGGCCACGTCAGATACACAAGCCACACAGAGAGACA 430

Qy  362 GATCAAAGGCAGCTCATAAAACCAAATGGTTTTAAAACATCACCCAGACAAAACGGAAAGCAGC 421
        |||||||||||||||||||| ||||||||||| || ||||||||||||||||||||||||||
Db  431 GATCAAAGCAGCTCACAAAACAATGGTCTTTGAAAACATCACCCAGACAAAAGCCGGAAAGCGGC 490

Qy  422 TCGGTGAACCAATAAAAGAAGGAGATAATGACTACTTCACTTGCAATAACTAAAGCTTATGA 481
        || ||||||||||| ||||||||| |||||||||||||||||| |||||||||||||||||
Db  491 TGGGGAACCCATCAAAGAAGGAGACAACGACTACTTCACTTGTATAACGAAAAGCTTATGA 550

Qy  482 AAATGTTATCTGATCCAGTG 500
        || |||| |||||||||||
Db  551 AAATGTTGTCTGATCCAGTG 569
```

```
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 112661)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (06-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   3  (bases 1 to 112661)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     --------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            --------- Summary Statistics
            Center project name: H_RG276003
            ---------
            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7/, or see http://genome.wustl.edu/gsc
            mailto:egreen@nhgri.nih.gov , or send

            SOURCE INFORMATION:
            Clone CTA-27603 is from a release of the human BAC library
            CITB-HS-A.  The library contains cloned DNA from human sperm.  See:
            Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
            Kim et al., Genomics 34:213-8 (1996).  The clone is available from
            Research Genetics, Inc. (http://www.resgen.com).
            VECTOR: pBeloBAC11
            Selection: chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            This clone contains STS SHGC-54619 (NID:g27344441).

            This clone contains STS (NID:g ).

            The actual start of this clone is at base position 112261 of CTA-27603;
            actual end is at base position 112261 of CTA-27603.

            This clone contains STS SHGC-54619 (NID:g27344441).
FEATURES             Location/Qualifiers
     source          1..112661
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q22-q31.1"
                     /clone_lib="CITB-HS-A"
                     /clone="CTA-27603"
     repeat_region   90..393
                     /rpt_family="Alu"
     repeat_region   1153..1454
                     /rpt_family="Alu"
     repeat_region   1658..1841
                     /rpt_family="Alu"
     repeat_region   1897..2055
                     /rpt_family="MER2_type"
```

```
     repeat_region   2067..2351
                     /rpt_family="MIR"
     repeat_region   2379..2471
                     /rpt_family="Alu"
     repeat_region   3065..3590
                     /rpt_family="MER21_g"
     repeat_region   3963..4268
                     /rpt_family="Alu"
     repeat_region   4296..4424
                     /rpt_family="Alu"
     repeat_region   5010..5304
                     /rpt_family="MIR"
     repeat_region   5738..5835
                     /rpt_family="Alu"
     misc_feature    5893..6124
                     /note="match to EST Aa836317 (NID:g2910636) od36f02.s1"
     repeat_region   7345..8321
                     /rpt_family="L1"
     repeat_region   8329..8628
                     /rpt_family="Alu"
     repeat_region   8803..8920
                     /rpt_family="Alu"
     repeat_region   9004..9131
                     /rpt_family="Alu"
     gene            <10095..42190
                     /gene="WUGSC:H_RG276003.1a"
     gene            <10095..42190
                     /gene="WUGSC:H_RG276003.1b"
     CDS             join(<10095..10200,12805..12995,16948..17023,27005..27103,
                     28075..28216,30197..30277,31096..31161,31965..32056,
                     32127..32248,32678..32827,34915..35003,35081..35150,
                     37745..37929,38671..38771,38888..38995,41658..41812,
                     42116..42190)
                     /gene="WUGSC:H_RG276003.1a"
                     /note="similar to murine cell cycle regulator MIDA1;
                     similar to A57591 (PID:g2137417); H_RG276003.1a"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC15864.1"
                     /db_xref="GI:3115346"
                     /translation="DVSRCAHRARPGAIMLLPSAADGRGTAITHALTSASTLCQVEP
                     VGRWFEAFVKRNRNASASFQELEDKELSESEDEELQLEEFPMLKTLDPKDWKNQD
                     HYAVLGLGHVRYKATQRQIKAAHKAMVLKHHPDKRKAAGEPIKEGDNDYFTCIYKAYE
                     MLSDPVKRRAFNSVDPTFDNSVPSKSEAKDNFEVFTPVFERNSRWSNKNVPKLGDM
                     NSSFEDVDIFYSFWYNFDSWREFSYLDEEEKERAECRDERRWIEKQNRATRAQRKEE
                     MNRIRTLVDNAYSCDPRIKFKFKEEEAKAKEAEAKRKEQEAKEKQRQAELEAA
                     RLAKEKEEEVRQQALLAKKEKDIQKRAIKKERQKLRNSCKTWNHFSDNEAERVKMME
                     EVEKLCDRLELASLQCLNETLTSCTKEVGKAAALEKQIEEINEQIRKEKEEAEAARMRQA
                     SKNTEKSTGGGNGSKNWSEDDLQLLIKAVNLPFAGTNSRWEVIANYMNIHSSGVKR
                     TAKDVIGRAKSLQKLDPHQKDDINKKAAFDKFKKEHGVVPQADNATPSEREGPYTDFT
                     PWTTEEQKLLEQALKTYPVNTEPERWEKIAEAVPGRTKKDCMKRYKELVEMVKAKKAAQ
                     EQVLNASRAKK"
     CDS             join(<10095..10200,12805..12995,16948..17023,27005..27103,
                     28075..28216,30197..30277,31096..31161,31965..32056,
                     32127..32248,32678..32827,34915..35003,35081..35150,
                     37745..37929,38671..38771,38888..38995,41658..41812,
                     42116..42190)
                     /gene="WUGSC:H_RG276003.1b"
                     /note="probable splice variant of H_RG276003.1a;
                     H_RG276003.1b"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC15865.1"
                     /db_xref="GI:3115347"
                     /translation="DVSRCAHRARPGAIMLLPSAADGRGTAITHALTSASTLCQVEP
                     VGRWFEAFVKRNRNASASFQELEDKELSESEDEELQLEEFPMLKTLDPKDWKNQD
                     HYAVLGLGHVRYKATQRQIKAAHKAMVLKHHPDKRKAAGEPIKEGDNDYFTCIYKAYE
                     MLSDPVKRRAFNSVDPTFDNSVPSKSEAKDNFEVFTPVFERNSRWSNKNVPKLGDM
                     NSSFEDVDIFYSFWYNFDSWREFSYLDEEEKERAECRDERRWIEKQNRATRAQRKEE
                     MNRIRTLVDNAYSCDPRIKFKFKEEEAKAKEAEAKRKEQEAKEKQRQAELEAA
                     RLAKEKEEEVRQQALLAKKEKDIQKRAIKKERQKLRNSCKIEEINEQIRKEKEEAEA
                     RMRQASKNTEKSTGGGNGSKNWSEDDLQLLIKAVNLPFAGTNSRWEVIANYMNIHSS
                     SGVKRTAKDVIGRAKSLQKLDPHQKDDINKKAAFDKFKKEHGVVPQADNATPSEREGP"
```

```
                   YTDFTPWTTEEQKLLEQALKTYPVNTPERWEKIAEAVPGFRKKDCMKRYKELVEMVKA
                   KKAAQEQVLNASRAKK"
repeat_region      11174..11238
                   /rpt_family="MIR"
repeat_region      11461..11756
                   /rpt_family="Alu"
repeat_region      11852..12148
                   /rpt_family="Alu"
repeat_region      12159..12456
                   /rpt_family="Alu"
repeat_region      13502..14021
                   /rpt_family="Alu"
repeat_region      14062..14303
                   /rpt_family="L2"
repeat_region      14296..14397
                   /rpt_family="MIR"
repeat_region      14622..14910
                   /rpt_family="Alu"
repeat_region      14985..15154
                   /rpt_family="Alu"
repeat_region      15155..15355
                   /rpt_family="Alu"
repeat_region      15471..15660
                   /rpt_family="L2"
repeat_region      15830..16129
                   /rpt_family="L2"
repeat_region      16161..16461
                   /rpt_family="Alu"
repeat_region      16491..16774
                   /rpt_family="Alu"
repeat_region      16814..16811
                   /rpt_family="Alu"
repeat_region      16583..18873
                   /rpt_family="Alu"
repeat_region      18916..19223
                   /rpt_family="Alu"
repeat_region      19627..19932
                   /rpt_family="Alu"
repeat_region      20011..20128
                   /rpt_family="Alu"
repeat_region      20166..20479
                   /rpt_family="Alu"
```

```
Query Match       38.6%;   Score 193.2;  DB 9;   Length 112661;
Best Local Similarity  98.5%;   Pred. No. 7.7e-41;
Matches  195;  Conservative  0;  Mismatches  3;  Indels  0;  Gaps  0;

Qy   102 ACCTCTGCCTCTACACTCTGTCTCAAGTTGAACCCTGTGGGAAGATGGTTTGAAGCTTTTGTT 161
             | | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 12798 ATCACAGCCTCTACACTCTGTCTCAAGTTGAACCCTGTGGGAAGATGGTTTGAAGCTTTTGTT 12857

Qy   162 AAGAGGAGAGAAACAGAAAATGCTTCTGCCTCTTTTTTCAGGAACCTGGAGGATAAGAAAAGAGGTTA 221
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 12858 AAGAGGAGAGAAACAGAAAATGCTTCTGCCTCTTTTTTCAGGAACCTGGAGGATAAGAAAAGAGGTTA 12917

Qy   222 TCCGAGGAATCAGAAGATGAAGAAATTGCAGTTGGAAGAGTTTCCCATGCTGAAAACACTT 281
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 12918 TCCGAGGAATCAGAAGATGAAGAAATTGCAGTTGGAAGAGTTTCCCATGCTGAAAACACTT 12977

Qy   282 GATCCCAAAAGGACTGGAAG 299
             ||||||||||||||||||||
Db 12978 GATCCCAAAAGGACTGGAAG 12995
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196130)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196130)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Dec 15, 2001 this sequence version replaced gi:17737069.
            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information --------------
            Center project name: H_NH1252L15
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 41%
            Sequencing vector: plasmid; 59%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192710 bases at least Q40
            Consensus quality: 193495 bases at least Q30
            Consensus quality: 195104 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 196979; sum-of-contigs
            Quality coverage: 8.79 in Q20 bases; agarose-fp
            Quality coverage: 8.63 in Q20 bases; sum-of-contigs
            ------------------------------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1    1194: contig of 1194 bp in length
          *   1195    1294: gap of unknown length
          *   1295    2762: contig of 1468 bp in length
          *   2763    2862: gap of unknown length
          *   2863   70873: contig of 68011 bp in length
          *  70874   70973: gap of unknown length
          *  70974  196130: contig of 125157 bp in length.
FEATURES             Location/Qualifiers
     source          1..196130
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-1252L15"
     misc_feature    1..1194
                     /note="assembly_name:Contig8"
     misc_feature    1295..2762
                     /note="assembly_name:Contig9"
     misc_feature    2863..70873
                     /note="assembly_name:Contig10"
     misc_feature    70974..196130
                     /note="assembly_name:Contig11
                     clone_end:SP6
                     vector_side:left"
BASE COUNT     56814 a  40018 c  37935 g  61063 t       300 others
ORIGIN
```

```
Query Match       38.6%;   Score 193.2;  DB 2;   Length 196130;
Best Local Similarity  98.5%;   Pred. No. 8.5e-41;
Matches  195;  Conservative  0;  Mismatches  3;  Indels  0;  Gaps  0;
```

```
Qy    102  ACCTCTGCCTCTACACTCTGTCAAGTTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTT  161
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157119  ATCACAGCCTCTACACTCTGTCAAGTTGAAACCTGTGGGAAGATGGTTTGAAGCTTTTGTT  157060

Qy    162  AAGAGGAGAAACAGAAAATGCTTCTGCCTCTTTTTCAGGAACTGAGGAGATAAGAAGAGTTA  221
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157059  AAGAGAGAAACAGAAAATGCTTCTGCCTCTTTTTCAGGAACTGAGGAGATAAGAAGAGTTA  157000

Qy    222  TCCGAGGAATCAGAAGATGAAGAATTGCAGTTGGAAGAGTTCCCATGCTGTGAAAAACACTT  281
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 156999  TCCGAGGAATCAGAAGATGAAGAATTGCAGTTGGAAGAGTTTCCCATGCTGTGAAAAACACTT  156940

Qy    282  GATCCCAAAAGACTGGAAG  299
           |||||||||||||||||||
Db 156939  GATCCCAAAAGACTGGAAG  156922
```

RESULT 8

```
LOCUS       AC023284             189242 bp    DNA     linear   HTG 11-FEB-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-255016 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION   AC023284
VERSION     AC023284.1  GI:6960385
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189242)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Bouffard,G.G.,
            Dietrich,N.L., Eagle,W.O., Gan,W., Gupta,J., Ho,S.-L., Huang,M.C.,
            Idol,J., Jamison,D.C., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
            Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
            Stantripop,S., Summers,T.J., Thomas,J.T.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189242)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
            Center project name: yn
            Center clone name: 255016
            -------------- Summary Statistics
            Sequencing vector: plasmid: n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 169396 bases at least Q40
            Consensus quality: 176085 bases at least Q30
            Consensus quality: 18018 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 228000; pulse-field-gel
            Quality coverage: 6.87x in Q20 bases; agarose-fp
            Quality coverage: 5.21x in Q20 bases; pulse-field-gel
            Quality coverage: 6.28x in Q20 bases; sum-of-contigs
            *
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..189242
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone_lib="RPCI mouse BAC library 23"
                     /clone="RP23-255016"
     *       1     2200: contig of 2200 bp in length
     *    2201     4687: gap of unknown length
     *    4688     7122: contig of 2435 bp in length
     *    7123    11237: gap of unknown length
     *   11238    14636: contig of 3399 bp in length
     *   14637    18498: gap of unknown length
     *   18499    24258: contig of 5760 bp in length
     *   24259    28714: gap of unknown length
     *   28715    32667: contig of 3953 bp in length
     *   32668    38867: gap of unknown length
     *   38868    42431: contig of 3564 bp in length
     *   42432    48750: gap of unknown length
     *   48751    54605: contig of 5855 bp in length
     *   54606    67106: gap of unknown length
     *   67107    73809: contig of 6703 bp in length
     *   73810    83974: gap of unknown length
     *   83975    94892: contig of 10918 bp in length
     *   94893   105454: gap of unknown length
     *  105455   117289: contig of 11835 bp in length
     *  117290   132852: gap of unknown length
     *  132853   148725: contig of 15873 bp in length
     *  148726   169036: gap of unknown length
     *  169037   189242: contig of 20206 bp in length
BASE COUNT    53124 a   40082 c   41371 g   54595 t   70 others
ORIGIN
```

```
Query Match  34.4%;  Score 172;  DB 2;  Length 189242;
Best Local Similarity  92.3%;
Matches 181;  Conservative 0;  Mismatches 15;  Indels 0;  Gaps 0;

Qy    104  CTCTGCCTCTACACTCTGTCAAGTTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTTA  163
           | || ||||| ||| |||||||||||||||||||||||||||||||||||||||||||||
Db 164338  CACAGCCTCTTCAGTCTGTCAAGTTGAACCTGTGGGAAGATGGTTTGAAGCTTTTGTTAA  164278

Qy    164  GAGGAGAAACAGAAAATGCTTCTGCCTCTTTTTCAGGAACTGAGGAGATAAGAAGAGTTATC  223
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db 164278  GAGGAGAAACAGAAATGCTTCTGCCTCTTTTTCAGGAACTGAGGAGATAAGAAGAGTTATC  164337

Qy    224  CGAGGAATCAGAAGATGAAGAATTGCAGTTGGAAGAGTTCCCATGCTGTGAAAAACACTTGA  283
           |||||||||||||||||||||||||||||| ||||| |||| ||||||||| |||||||||||
Db 164398  AGAGGAATCAGAAGATGAAGAATTGCAGTTACAGTTAGAAGAATTCCCATGCTGAAAAACACTTGA  164457
```

```
Qy      284 TCCCAAGAGACTGGAAG 299
            ||||||||| ||||||
Db 164458 TCCCAAGAGATTGGAAG 164473
```

```
RESULT 9
LOCUS       HS217P22              134550 bp    DNA      linear   PRI 23-NOV-1999
DEFINITION  HS217P22/c
            Human DNA sequence from clone 217P22 on chromosome 6p21.1-21.31.
            Contains part of a novel Dynein heavy chain family protein gene and
            a pseudogene similar to mouse MIDA1. Contains an EST, an STS and
            GSSs, complete sequence.
ACCESSION   AL035555
VERSION     AL035555.10  GI:4581305
KEYWORDS    HTG; Dynein heavy chain; MIDA1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134550)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 12, 1999 this sequence version replaced gi:4490837.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 217P22. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            217P22 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..134550
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p21.1-21.31"
                     /clone="RP1-217P22"
                     /clone_lib="RPCI-1"
     mRNA            1..62
                     /gene="dJ271P22.1"
     repeat_region   90..498
                     /note="L1MD2 repeat: matches 5893..6341 of consensus"
     repeat_region   complement(join(<649..756,848..933,3464..3651,5066..5244,
                     6895..7067,9782..10006,15850..15952,17110..17273,
                     18723..18876,21751..21950,24621..24861,25019..25156,
                     29482..29582,32883..32945,35089..35202,39236..39340,
                     41133..41368,43836..43962,44490..44686,51867..>52056))
                     /gene="dJ271P22.1"
                     /note="match: cDNAs AJ132091 AJ132090 283806 283817 D26499
                     D26496
                     match: proteins Q39575 P45443 P78716 Q39610 P37276 Q34036
                     P23098 Q27171 Q39057 Q19020 CE13884 Q92815 Q38650 Q39565
                     P45444 Q77075
                     this gene fragment and fragments dJ271P22.1a and
                     dJ271P22.1b are parts of one gene; the exon structure
                     linking them is yet to be determined"
     gene            <105721..>105822,111445..111552,116935..117048,
                     125637..>125813)
                     /gene="dJ271P22.1"
     repeat_region   1415..1480
                     /note="3 copies 22 mer 86% conserved"
     repeat_region   1651..1702
                     /note="L2 repeat: matches 2641..2692 of consensus"
     repeat_region   1703..1836
                     /note="MLT1J repeat: matches 6..146 of consensus"
     repeat_region   2645..2766
                     /note="L2 repeat: matches 1804..2003 of consensus"
     repeat_region   2770..3164
                     /note="MIR repeat: matches 66..205 of consensus"
     repeat_region   3870..4072
                     /note="MLT1I repeat: matches 28..410 of consensus"
     repeat_region   4073..4369
                     /note="AluSg repeat: matches 1..307 of consensus"
     repeat_region   4370..4595
                     /note="L2 repeat: matches 1569..1804 of consensus"
     repeat_region   5519..5692
                     /note="MER63A repeat: matches 1..184 of consensus"
     repeat_region   5693..5964
                     /note="AluSq repeat: matches 1..296 of consensus"
     misc_feature    5965..6012
                     /note="MER63A repeat: matches 184..210 of consensus"
     misc_feature    6164..6572
                     /note="match: STS AL035191"
     repeat_region   7173..7694
                     /note="L2 repeat: matches 2158..2700 of consensus"
     misc_feature    complement(8093..8440)
                     /gene="dJ271P22.1"
                     /note="match: EST AA701401"
     repeat_region   8837..9001
                     /note="MER5A repeat: matches 4..183 of consensus"
     misc_feature    10525..11011
                     /note="match: GSS B14808"
     repeat_region   11315..11487
                     /note="L1MA5A repeat: matches 5906..6079 of consensus"
     repeat_region   11489..11804
                     /note="MER1B repeat: matches 1..337 of consensus"
     repeat_region   11805..12018
                     /note="L1MA5A repeat: matches 6073..6293 of consensus"
     repeat_region   12145..12252
                     /note="L1PA9 repeat: matches 6055..6163 of consensus"
     repeat_region   12426..12531
                     /note="53 copies 2 mer at 74% conserved"
     repeat_region   12445..12532
                     /note="4 copies 22 mer 77% conserved"
     repeat_region   12657..12700
                     /note="22 copies 2 mer at 75% conserved"
     repeat_region   13447..13484
                     /note="19 copies 2 mer at 79% conserved"
     repeat_region   13512..13876
                     /note="MER47A repeat: matches 1..366 of consensus"
     repeat_region   15463..15652
                     /note="MIR repeat: matches 61..261 of consensus"
     repeat_region   16252..16371
                     /note="MER94 repeat: matches 2..134 of consensus"
     repeat_region   16518..16923
                     /note="MLT1C repeat: matches 6..466 of consensus"
     repeat_region   18402..18616
                     /note="MER58A repeat: matches 3..224 of consensus"
     misc_feature    complement(19014..19463)
```

a frameshift around 895 disrupts the ORF of this gene,
compared to all matching sequences; there are no
indications of errors in our sequence
supported by FGENES and GENSCAN
some exon boundaries derived from FGENES and GENSCAN
predictions"
/evidence=not_experimental
/product="dJ271P22.1 (novel Dynein heavy chain family
protein)"
complement(join(649..86255,89060..89136,91573..91770,

```
misc_feature    /gene="dJ271P22.1"
                /note="match: GSS AQ235524"
repeat_region   19463..19966
                /note="match: GSSs AQ244081 B81565"
repeat_region   23464..23559
                /note="L1M1 repeat: matches 253..346 of consensus"
repeat_region   23921..23972
                /note="L1MC5 repeat: matches 7805..7856 of consensus"
repeat_region   24157..24300
                /note="L1M2 repeat: matches 1173..1330 of consensus"
repeat_region   25643..25798
                /note="Looper repeat: matches 1..157 of consensus"
repeat_region   26080..26387
                /note="FRAM repeat: matches 1..308 of consensus"
repeat_region   26451..27564
                /note="AluSx repeat: matches 1..1255 of consensus"
repeat_region   27565..27857
                /note="Looper repeat: matches 159..1255 of consensus"
repeat_region   27867..28157
                /note="AluSp repeat: matches 3..296 of consensus"
repeat_region   28158..28282
                /note="AluJb repeat: matches 1..301 of consensus"
repeat_region   28302..28433
                /note="L2 repeat: matches 1262..1427 of consensus"
repeat_region   28324..28468
                /note="MER5A repeat: matches 57..189 of consensus"
repeat_region   28826..29120
                /note="MER5A repeat: matches 35..189 of consensus"
repeat_region   3093..31038
                /note="AluY repeat: matches 1..294 of consensus"
repeat_region   31048..31125
                /note="23 copies 2 mer tt 76% conserved"
repeat_region   31356..32006
                /note="AluSg/x repeat: matches 133..310 of consensus"
repeat_region   32246..32532
                /note="L2 repeat: matches 1814..2488 of consensus"
repeat_region   33102..33226
                /note="L2 repeat: matches 485..674 of consensus"
repeat_region   33692..33976
                /note="L1MB8 repeat: matches 6044..6171 of consensus"
repeat_region   3443..35263
                /note="MER93 repeat: matches 91..395 of consensus"
misc_feature    34711..34931
                /note="match: GSS AQ241294"
repeat_region   35667..35912
                /note="MER30 repeat: matches 1..230 of consensus"
repeat_region   36004..36381
                /note="MIR repeat: matches 13..262 of consensus"
repeat_region   36867..37171
                /note="AluJb repeat: matches 1..312 of consensus"
repeat_region   37229..37567
                /note="AluY repeat: matches 1..304 of consensus"
repeat_region   37566..37617
                /note="L1M4 repeat: matches 2988..3343 of consensus"
repeat_region   37605..37758
                /note="L1P repeat: matches 5176..5225 of consensus"
repeat_region   37767..38455
                /note="L1MB2 repeat: matches 5991..6155 of consensus"
repeat_region   38455..38804
                /note="L1M4 repeat: matches 4731..5396 of consensus"
repeat_region   38817..39147
                /note="L1MB7 repeat: matches 5821..6173 of consensus"
                /note="MER7A repeat: matches 1..346 of consensus"
```

```
Query Match 33.5%;  Score 167.4;  DB 9;  Length 134550;
Best Local Similarity 84.7%;  Pred. No. 6.9e-34;
Matches 216;  Conservative 0;  Mismatches 26;  Indels 13;  Gaps 2;

Qy    219 TTATCCGAGGAATCAGAAGATGAAGAATTGCAGTTGGAAGAGTTTCCCATGCTGAAAACA 278
          |||  ||  |||| |||||||||||||||||||||||||||||||||||
Db 104634 TGAACTGGAGGATAAGAAGAGAAGAGAATTGCAGTT---AGTGTTTCTCATGCTGAAAACA 104578
Qy    279 CTTGATCCCAAAGACTGGAAGAACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTA 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104577 GTTGATCCCAAGGACTGGAAGAATCAGGATGAATCAGGATTCATTATGCAGTTCTTGCACTTAGCCAGTGA 104518
Qy    339 AGATACAAGGCTACACAGAGACGACAGATCAAAGAGAGCTCATAAAGCAATGGTGTTTTAAAAACAT 398
          ||||||||| |||||||||||| |||||||| |||||||||| |||| || ||| ||||||||||||
Db 104517 AGATACAAAGGCTACACAGAGACAGATCAAAAGCAGCTCATTAAAGTGAGTTGGTTTTAAAAACAT 104458
Qy    399 CACCCAGACAAACGGAAAGCAGCTGTGGTGAACCAATAAAAAGAGGAGGAGATAATGACTACTTC 458
          |          ||||||||||| |||| |||||||||||||| ||||  ||||||||||||||||
Db 104457 C------------ATGGAAAGCAGCCAGTGAACCAATAAAAAGAAAAAGGGAATAATGACTACTTC 104408
Qy    459 ACTTGCATAAACTAA 473
          ||||||||| ||||
Db 104407 ACTTGCATCACTGAA 104393
```

```
RESULT 10
AC022402
LOCUS       AC022402              164268 bp    DNA     linear   PRI 25-JUN-2000
DEFINITION  Homo sapiens chromosome 06 clone RP11-7013, complete sequence.
ACCESSION   AC022402
VERSION     AC022402.4  GI:8705033
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164268)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164268)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 164268)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Jun 25, 2000 this sequence version replaced gi:8569085.
FEATURES             Location/Qualifiers
     source          1..164268
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="06"
                     /clone="RP11-7013"
                     /clone_lib="RPCI-11"
BASE COUNT    48064 a   31200 c   31866 g   53138 t
ORIGIN
```

```
Query Match 33.5%;  Score 167.4;  DB 9;  Length 164268;
Best Local Similarity 84.7%;  Pred. No. 7.1e-34;
Matches 216;  Conservative 0;  Mismatches 26;  Indels 13;  Gaps 2;

Qy    219 TTATCCGAGGAATCAGAAGATGAAGAATTGCAGTTGGAAGAGTTTCCCATGCTGAAAACA 278
          |||  ||  |||| |||||||||||||||||||||||||||||||||||
Db   6950 TGAACTGGAGGATAAGAAGAGGAAGAGAATTGCAGTT---AGTGTTTCTCATGCTGAAAACA 7006
Qy    279 CTTGATCCCAAAGACTGGAAGAACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTA 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7007 GTTGATCCCAAGGACTGGAAGAATCAGGATTCATTATGCAGTTCTTGGACTTAGCCAGTGA 7066
Qy    339 AGATACAAGGCTACACAGAGACGACAGATCAAAGAGAGCTCATAAAGCAATGGTGTTTTAAAAACAT 398
          ||||||||| |||||||||||| |||||||| |||||||||| |||| || ||| ||||||||||||
Db   7067 AGATACAAAGGCTACACAGAGACAGATCAAAAGCAGCTCATTAAAGTGAGTTGGTTTTAAAAACAT 7126
Qy    399 CACCCAGACAAACGGAAAGCAGCTGTGGTGAACCAATAAAAAGAGGAGGAGATAATGACTACTTC 458
          |          ||||||||||| |||| |||||||||||||| ||||  ||||||||||||||||
Db   7127 C------------ATGGAAAGCAGCCAGTGAACCAATAAAAAGAAAAAGGGAATAATGACTACTTC 7176
Qy    459 ACTTGCATAAACTAA 473
          ||||||||| ||||
Db   7007 ACTTGCATCACTGAA 473
```

```
Db  7177  ACTTGCATCACTGAA 7191
          ||||||||| ||| ||
          ||||||||| ||| ||
Qy  7177  ACTTGCATCACTGAA 7191
```

RESULT 11

```
LOCUS       AC023284            192242 bp    DNA     linear   HTG 11-FEB-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-255016 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION   AC023284
VERSION     AC023284.1  GI:6960385
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189242)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Bouffard,G.G.,
            Dietrich,N.L., Eagle,W.O., Gan,W., Gupta,J., Ho,S.-L., Huang,M.C.,
            Idol,J., Jamison,D.C., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
            Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
            Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189242)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information --------------
            Center project name: yn
            Center clone name: 255016
            -------------- Summary Statistics --------------
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 169396 bases at least Q40
            Consensus quality: 176085 bases at least Q30
            Consensus quality: 180318 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 228000; pulse-field-gel
            Insert size: 189242; sum-of-contigs
            Quality coverage: 6.87x in Q20 bases; agarose-fp
            Quality coverage: 5.21x in Q20 bases; pulse-field-gel
            Quality coverage: 6.28x in Q20 bases; sum-of-contigs
            ------------------------------------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 23 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *      1      2200: contig of 2200 bp in length
          *   2201      4687: gap of unknown length
          *   4688      7122: contig of 2487 bp in length
          *   7123     11237: gap of unknown length
          *  11238     14636: contig of 2435 bp in length
          *  14637     18498: gap of unknown length
          *                   contig of 4415 bp in length
          *                   gap of unknown length
          *                   contig of 3399 bp in length
          *                   gap of unknown length
          *                   contig of 3862 bp in length
```

```
                                             18499     24258: contig of 5760 bp in length
                                             24259     28714: gap of unknown length
                                             28715     32667: contig of 4456 bp in length
                                             32668     38867: contig of 3953 bp in length
                                             38868     42431: gap of unknown length
                                             42432     48750: contig of 6200 bp in length
                                             48751     54605: gap of unknown length
                                             54606     67106: contig of 3564 bp in length
                                             67107     73809: gap of unknown length
                                             73810     83974: gap of unknown length
                                             83975     94892: contig of 12501 bp in length
                                             94893    105454: gap of unknown length
                                            105455    117289: contig of 6703 bp in length
                                            117290    132852: gap of unknown length
                                            132853    148725: contig of 10165 bp in length
                                            148726    169036: gap of unknown length
                                            169037    189242: contig of 10918 bp in length
                                                               gap of unknown length
                                                               contig of 10562 bp in length
                                                               gap of unknown length
                                                               contig of 11835 bp in length
                                                               gap of unknown length
                                                               contig of 15563 bp in length
                                                               gap of unknown length
                                                               contig of 15873 bp in length
                                                               gap of unknown length
                                                               contig of 20311 bp in length
                                                               gap of unknown length
                                                               contig of 20206 bp in length.
FEATURES             Location/Qualifiers
     source          1..189242
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone_lib="RPCI mouse BAC library 23"
                     /clone="RP23-255016"
BASE COUNT    53124 a   40082 c   41371 g   54595 t   70 others
ORIGIN
```

```
Query Match   18.6%;  Score 93;  DB 2;  Length 189242;
Best Local Similarity   84.0%;  Pred. No. 7.3e-14;
Matches 105;  Conservative 0;  Mismatches 20;  Indels 0;  Gaps 0;

Qy   360 CAGATCAAAGCAGCTCATAAAGCAAGGTGTTTTAAAACATCACCCAGACACAAACGGAAACGA 419
         ||| || |||  |||| ||| ||||||| ||||||||||| |||| |||||||||||| ||||
Db 93288 CTGATTGATCCTGCAGATAAAGCAATGGTCTTGAAACATCATCCAGACAGAACGGAAAGCG 93229

Qy   420 GCTGGTGAACCAATAAAAGAAGGAGAGATAATGAACTACTTCACTTGCATAACTAAAGCTTAT 479
         ||||| ||||| || ||| ||||||| || |||||| ||||||||||| |||||||||||| |
Db 93228 GCTGGGGAACCCATCAAAGAAGGAGAGCAATGACTACTTCACTTGCATAACTAAAGGTTAAG 93169

Qy   480 GAAAT 484
         | |||
Db 93168 AAAAT 93164
```

RESULT 12

```
LOCUS       AY069523            2248 bp    mRNA    linear   INV 17-DEC-2001
DEFINITION  Drosophila melanogaster LD23875 full length cDNA.
ACCESSION   AY069523
VERSION     AY069523.1  GI:17862381
KEYWORDS    FLI_cDNA.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
```

REFERENCE　1 (bases 1 to 2248)
AUTHORS　Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
　　　　Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
　　　　Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
　　　　Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K.,
　　　　Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.

TITLE　　Direct Submission
JOURNAL　Submitted (10-DEC-2001) Berkeley Drosophila Genome Project,
　　　　Lawrence Berkeley National Laboratory, One Cyclotron Road,
　　　　Berkeley, CA 94720, USA

COMMENT　Sequence submitted by:
　　　　Berkeley Drosophila Genome Project
　　　　Lawrence Berkeley National Laboratory
　　　　Berkeley, CA 94720
　　　　This clone was sequenced as part of a high-throughput process to
　　　　sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
　　　　Science 2000). The sequence has been subjected to integrity checks
　　　　for sequence accuracy, presence of a polyA tail and contiguity
　　　　within 100 kb in the genome. Thus we believe the sequence to
　　　　reflect accurately this particular cDNA clone. However, there are
　　　　artifacts associated with the generation of cDNA clones that may
　　　　have not been detected in our initial analyses such as internal
　　　　priming from contaminating genomic DNA, retained introns
　　　　due to reverse transcription of unspliced precursor RNAs, and
　　　　reverse transcriptase errors that result in single base changes.
　　　　For further information about this sequence, including its location
　　　　and relationship to other sequences, please visit our Web site
　　　　(http://fruitfly.berkeley.edu) or send email to
　　　　cdna@fruitfly.berkeley.edu.

FEATURES　　　　Location/Qualifiers
　　source　1..2248
　　　　/organism="Drosophila melanogaster"
　　　　/strain="y; cn bw sp"
　　　　/db_xref="taxon:7227"
　　　　/map="78B4-78C1"
　　gene　1..2248
　　　　/gene="CG10565"
　　　　/note="alignment with genomic scaffold AE003593"
　　　　/db_xref="FLYBASE:FBgn0037051"
　　CDS　151..2091
　　　　/gene="CG10565"
　　　　/note="longest ORF"
　　　　/codon_start=1
　　　　/db_xref="FLYBASE:FBgn0037051"
　　　　/product="LD23875p"
　　　　/protein_id="AAL39668.1"
　　　　/db_xref="GI:17862382"
　　　　/translation="MTSGTVATAVEVQLPLKVVRRKIERVGFAYAFAQRROFLAPGGVE
　　　　RSESDEKLEGVEGEEVDISYLKSLDDYHAVLGLGKLRYEASEDDVRRAYRRM
　　　　VLLHPDRKAKGEEVIQDDDYFPCITKAYEILGTSKPRSFDSVDPEFDSLPSQND
　　　　LDNDYFGVFNKFFTLNGRWSEKPHVPSFGGVDAKREEVERYFNFWDFKSWREFSYLD
　　　　EEDKEKGQDRDERWIEKENRARIRKRKEEMSIRTSLVDLAYNNDKRIORFKQEKD
　　　　RKAAKRAKMDAAQAOKAEADRAIREAALAKEKAEKAEQKRIEQTRIEREQOKKLLKK
　　　　ERTLRDKVKDCKYYAKNDKDLKHMEGTEKICETFNLAELQALNKAMESKGRESFVA
　　　　ALQTAEORIAAELEEINQTQAAKKLASSAATPKGYKVEKVKNELWSNENVOLLIKAVNLF
　　　　PAGTAQORDVIATFINQHSPDNTVLVNARDVLNRAKALQNTDHSRKSSLKTQANDAAFA
　　　　SFEKSKKDVQTCKDITLGEETAQASKENLKQNGVDHKANNQSTKQNGTAPAPANPTAA
　　　　PAPVPATNGSTGGGAASKTWTKEQALLEQAIKIYPTTPDRWDCIAACIPNRSKKDC
　　　　LRRVKELVELVNSKKEAQAAVK"

BASE COUNT　656 a　555 c　630 g　407 t
ORIGIN

　Qy　224　CGAGGAATCAGAGAGAGAATGAGAAATTGGAAGAAGAGAGTTTCCCATCGTGAAAACACTTGA　283
　　　　　|| || | ||||| ||| | | | || |||| |||| || | |||||| ||| | ||
　Db　294　CGATGAGAAATTGGAGGGCGTGGGCGCCGAGGAGGTGGACATCAGCTACTCACTAAAGTCGCTGGA　353

　Qy　284　TCCCAAAAGCTGGAAGAGAACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGATA　343
　　　　　|| ||| | |||||| | | | | ||||| | |||| ||| | |||| | ||||
　Db　354　TCCTAAGGAGTGGAGAGACCAGAGCATTACGCCGTTCTTCGGCTTGGGCCAAGCTCCGATA　413

　Qy　344　CAAGGCTACACAGAGACAGATCAAAGCAGCTCATAAAGCAATGGTTGTTTTAAAACATCACC　403
　　　　　| ||||| || || | || | || |||| |||| | ||||| || ||||| |
　Db　414　CGAGGCCAGCGAGGATGAGTGTTCGACGGGCCTTACAGCGCATTACGGAGATAATGACTACTTCACTTG　473

　Qy　404　AGACAAACGCGAAAGCAGCTGTGACCAATAAAAGAGGAGGAGATAATGACTACTTCACTTG　463
　　　　　|| || |||||| ||||| | || |||||| || | | | ||||||| || ||||
　Db　474　CGATAAGCGGAAAGCCAAGGGCGAGGAAGTCATCCA---GGACGATGAGATTACTTCACATG　530

　Qy　464　CATAACTAAAGCTTTATGAAAT　484
　　　　　||||| |||||||| || | |
　Db　531　CATAACCAAAGCTTACGAGAT　551

RESULT 13
LOCUS　SPBC30D10/c
　　　　SPBC30D10　　37838 bp　　DNA　　　linear　　PLN 03-NOV-1999
DEFINITION　S.pombe chromosome II cosmid c30D10.
ACCESSION　Z97992
VERSION　Z97992.2　GI:6249531
KEYWORDS　60s ribosomal protein; 60s ribosomal protein L10; biotin-protein
　　　　ligase; chorismate mutase; dnaJ domain; glucan synthase-1; glycine
　　　　rich; gpi; Homol D box; mitochondrial genome maintenance;
　　　　N-acetylglucosaminyl-phosphatidylinositol; oxidoreductase; pdb1;
　　　　phosphatidylinositol 3-kinase; pyruvate dehydrogenase e1 component;
　　　　rpl10; serine protease; short chain dehydrogenase; small
　　　　ribonuclear protein; transcriptional regulator; zdha binding;
　　　　zuotin like.
SOURCE　fission yeast.
　ORGANISM　Schizosaccharomyces pombe
　　　　Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
　　　　Schizosaccharomycetales; Schizosaccharomycetaceae;
　　　　Schizosaccharomyces.
REFERENCE　1 (bases 1 to 29767)
AUTHORS　Wood,V., Rajandream,M.A., Barrell,B.G., Lauber,J., Hilbert,H. and
　　　　Duesterhoeft,A.
TITLE　　Direct Submission
JOURNAL　Submitted (02-FEB-1998) European Schizosaccharomyces genome
　　　　sequencing project, Sanger Centre, The Wellcome Trust Genome
　　　　Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
　　　　and QIAGEN GmbH, Max-Volmer-Str 4, D-40724 Hilden, Germany
COMMENT　On Nov 4, 1999 this sequence version replaced gi:2276349.
　　　　Notes:
　　　　Details of yeast sequencing at the Sanger Centre are available on
　　　　the World Wide Web.
　　　　(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
　　　　During 1995 to 1996 about 80% of S. pombe chromosome 1 was
　　　　sequenced by the Sanger Centre. The sequencing of the S. pombe
　　　　genome is now being continued with funding from The European
　　　　Commission. Fourteen European sequencing laboratories, including
　　　　the Sanger Centre, are participating in the project.
　　　　Protein coding regions (CDS) have been predicted with the help of
　　　　computer analysis using the Genefinder program in PomBase (an ACEDB
　　　　database) with additional predictions for the branch-acceptor sites
　　　　supplied by the program Sp3splice. CAUTION: It is possible that for
　　　　any individual CDS we may have underestimated or overestimated the
　　　　number of introns/exons or we may not have chosen the correct
　　　　splice donor/acceptor sites.
　　　　CDS are numbered using the following system eg SPBC25H2.01c (S.
　　　　pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
　　　　(complementary strand).
　　　　The more significant matches with motifs in the PROSITE database
　　　　are also included but some of these may be fortuitous.
　　　　The length in codons is given for each CDS.
　　　　IMPORTANT: This sequence MAY NOT be the entire insert of the
　　　　sequenced clone. It may be shorter because we only sequence
　　　　overlapping sections once, or longer, because we arrange for a
　　　　small overlap between neighbouring submissions.
　　　　Cosmid c30D10 does not apply.
FEATURES　　　　Location/Qualifiers
　　source　1..37838
　　　　/organism="Schizosaccharomyces pombe"
　　　　/strain="972h-"

misc_feature
/db_xref="taxon:4896"
/chromosome="II"
/map="IIR"
/clone="cosmid c30D10"
1..1450

misc_feature
/note="overlap with c13E7 S. pombe chromosome 2"

misc_feature
complement(8..13)
/note="gtaagt, splice donor sequence"

gene
complement(join(14..110,172..236))
/gene="SPBC30D10.20"

CDS
complement(join(14..110,172..236))
/gene="SPBC13E7.10c"
/note="SPBC13E7.10c"
/codon_start=1
/partial
/product="transcription factor IIIb subunit"
/protein_id="CAB60095.1"
/db_xref="GI:6249532"
/translation="MGCPNCGSTTFESDTASGNTYCTQCGVVVEQDAIVSEVIFGEAS
TGAAVVQGSL"

gene
/note="SPBC30D10.20"
912..1808

misc_feature
complement(111..125)
/gene="SPBC30D10.20"
/note="ctaacattgaatag, splice branch and acceptor"

misc_feature
complement(166..171)
/gene="SPBC30D10.20"
/note="gtaagt, splice donor sequence"

gene
912..1808
/gene="SPBC30D10.19c"

CDS
912..1808
/gene="SPBC30D10.19c"
/note="SPBC30D10.19c, len:298, SIMILARITY:Saccharomyces
cerevisiae, YG2Q_YEAST, hypothetical 38.8 kd protein ,
(346 aa), fasta scores: opt: 392, E():2.4e-19, (33.1%
identity in 260 aa)"
/codon_start=1
/label=SPBC30D10.19c
/product="hypothetical protein"
/protein_id="CAB10814.1"
/db_xref="GI:2276368"
/db_xref="SWISS-PROT:O14364"
/translation="MNRLNFQSLLMKFGQKRFYRPWIRIENIKPTLWKPITPFAVGV
GSATFYTANYLDKRKNYPKSSYGPFIPQTSRSLVLSIIGINVGVFALWRAPRFSHL
NRFLQKYAVMNPIFINMPSMIVSAFSHQSCWHLLFNMVAAFYSFAPAIVDVFGNNQFVA
FYISSLLFPSNVASLLHHRLRFGTKVTPGSLGASGAIAYAIAAATSYFFPNASVSIIFLP
FIPIKIGVALLGLMAFDAWGLISRGFSSFANFTLIDHAAHLGGGIFGWLYAKYGYSTY
NRSTRPRPPSLSKPFFSRVSF"

promoter
2272..2279
/note="Homol D box"

gene
join(2350..2737,2999..3261)
/gene="SPBC30D10.18c"

gene
/note="rpl1-2; rpl10a-2"
join(2350..2737,2999..3261)
/gene="SPBC30D10.18c"

CDS
join(2350..2737,2999..3261)
/gene="rpl10a-2"
/gene="SPBC30D10.18c"
/codon_start=1
/note="SPBC30D10.18c, len:216"
/product="60s ribosomal protein L10"
/protein_id="CAB10813.1"
/db_xref="GI:2276367"
/db_xref="SWISS-PROT:O14363"
/translation="MSKVSPANIRSSVETILKGSEBKKRNFTETVELQIGLKNYDPQR
DKRFSGTIKLPNYPRPNMSICILGDAHDILDRAKHGVDAMSAVDDLKKLNKRKLVKKL
AKKYDAFIASEVLIKQIPRLLGPGLSKAGKFPPSPVSHSDDLYGKIIEVKSTIKFQLKK
VLCLGVAVGHVDMAEEQLAANLSLAINFLVSLLKKGWQNIGSLVIKSTMGRPYRLY"

misc_feature
join(2359..2737,2999..3246)
/gene="SPBC30D10.18c"
/note="Match to PF00687 Ribosomal_L1, L1P family of
ribosomal proteins Score 217.61"

intron
2738..2998
/gene="SPBC30D10.18c"
/note="gtaagt, splice donor sequence"
/note="confirmed"

misc_feature
2981..2998
/gene="SPBC30D10.18c"
/note="ctaactcgatggctatag, splice branch and acceptor"

gene
4093..5607
/gene="SPBC30D10.17c"

CDS
4093..5607
/gene="SPBC30D10.17c"
/note="SPBC30D10.17c, len:504, SIMILARITY:Neurospora
crassa., GS1_NEUCR, glucan synthase-1, (532 aa), fasta
scores: opt: 523, E():4.7e-25, (36.2% identity in 549 aa)"
/codon_start=1
/label=SPBC30D10.17c
/product="putative glucan synthase"
/protein_id="CAB10812.1"
/db_xref="GI:2276366"
/db_xref="SPTREMBL:O14362"
/translation="MSKNSFSSMANSVTSFFQSLTTPNRHADPSFRPSRREKQSRLPT
PLQSVAASAYSGVNASQTGLLNDSRANSVTNLPNSSNTSQVGLNNISPAPVGVPGSK
TNELANNSMEMQEISPNGSASLPPPVSESWRRIDRWAEENYELYCQLCYGAATVADVD
SLYELECTLPRDVRESLYIHDGQDRGGQPPGILFGVTLLLDIEEIEESELWRRVAQS
YAEATLAGKIDQAVASRQASPPGAVQCVYAHPGWIPLAKDVGNNIAIDLAPGPAGQ
WGQVILFGRDQDTKYVVARSWADFLAIVAIDMENGKWNLVDEDDNSLRLIYGPREQWS
YLDILKYRARKAERRKFKKRDGKRTPIRKSIAKEDVTNSANSTAPSTSGTTVLDDGL
DNNVEDIPLYGPSKDDELIKKEELEADTDLGLINTSEIINQPANLPDEPTAETSNPVSA
TTVEAVTTTADNKDEEKNDHVTEDVSQNSTIAEASSLQAQEEREKEIETTSVKQE"

gene
5245..5277
/gene="SPBC30D10.17c"
/note="PS00136 Serine proteases, subtilase family,
aspartic acid active site"

misc_feature
5245..5277
/gene="SPBC30D10.16"

gene
complement(join(6003..6812,6958..7011))
/gene="SPBC30D10.16"

CDS
complement(join(6003..6812,6958..7011))
/gene="SPBC30D10.16"
/note="SPBC30D10.16, len:287, SIMILARITY:Neisseria
gonorrhoeae., Q9ZHY3, phea., (375 aa), fasta scores: opt:
418, E():8.5e-21, (32.1% identity in 290 aa)"
/codon_start=1
/label=SPBC30D10.16
/note="splicing may be incorrectly predicted"
/product="putative chorismate mutase"
/protein_id="CAB10811.1"
/db_xref="GI:2276365"
/db_xref="SPTREMBL:O14361"
/translation="MSASKIAFLPRGTFSHQAALLARPDSLLCSLPSFAGKLIFALK
FILICFLAVLEALSSRQVDAVLPIENSTNGAVIPAYDLLKGRDDIQAVGEVLVPAHH
CIIGKSLENVOKILSHPQAFGQCSKWISANVPNAEFVSVSSTSQAAALASKDITGTIV
AISSELCAVENQFNLLVKNIEDDSNNRTRLLLRSGGFQODDLSPLKEKSLLQFYLSHP
KKLSAVFEVFAAHKVVITNLVVRPSCKFPWTYIFVECLGMEKHLIDRVGKYCDTFFF
MGSYTNQISYF"

misc_feature
complement(join(6027..6757,6954..6996))
/gene="SPBC30D10.16"
/note="Match to PF00800 PDT, Prephenate dehydratase Score
123.61"

misc_feature
complement(6813..6828)
/gene="SPBC30D10.16"
/note="gtaacaaatataatag, splice branch and acceptor"

gene
complement(6952..6957)
/gene="SPBC30D10.16"
/note="gtgtgt, splice donor sequence"

CDS
complement(7727..9277)
/gene="SPBC30D10.15"
/note="SPBC30D10.15, len:516, SIMILARITY:Saccharomyces
cerevisiae, YNM4_YEAST, hypothetical 54.9 kd protein in

Query Match   14.6%;   Score 73.2; DB 8;   Length 37838;
Best Local Similarity 62.4%;   Pred. No. 1.2e-08;

Matches 136; Conservative 0; Mismatches 73; Indels 9; Gaps 1;

```
Qy    282 GATCCCAAAGACTGGGAGAAACCAAGATCATTATGCAGTTCTTGGACTTGGCCATGTGAGA 341
          ||||||| ||||||||||||||||| ||||||||||||||||||||||||| |||||||
Db  37838 GATCCCAAGGAATGGAAGACAACAAGATCATTATGCGTCTTAGGTCTTCCCAAGTACCGC 37779

Qy    342 TACAAGGCTACACAGACAGACAATCAAAAGCAGCTCATAAAAGCAATGGTTTTAAAACATCAC 401
          |||||       |||||||||||||||||||||||||||||||
Db  37778 TACAAAGCTGATACTGAGCAAATCAAGAAGAAGAAGCTCATTTGAAGAAGCATCAT 37719

Qy    402 CCAGAGAAAACGGAAAGCAGCTGGTGAACCAAATAAAAGAGGAGATAATGACTACTTCACT 461
          ||        ||| |||||||||||||||||||
Db  37718 CCTGATAAGAAAGCTGCCTCTGGT--------AACATAAACGATGATTCTTCTTCAAG 37668

Qy    462 TGCATAACTAAAGCTTATGAAATGTTATCTGATCCAGT 499
          ||||| |||||||||||||||||||||||||||||||
Db  37667 TGCATCCAAAAGGCTTATGAAATTTTATCTGATCCTGT 37630
```

RESULT 14
LOCUS   AL593452
AL593452
DEFINITION Meleagris gallopavo DNA; STS TC1078, sequence tagged site.
ACCESSION   AL593452
VERSION   AL593452.1   GI:14716570
KEYWORDS   STS.
SOURCE   turkey.
ORGANISM   Meleagris gallopavo
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Meleagrididae;
            Meleagris.
REFERENCE   1 (bases 1 to 680)
AUTHORS   Morrice,D.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 680)
AUTHORS   Morrice,D.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Morrice D., Roslin Institute, Roslin,
            Midlothian, EH25 9PS, UK david.morrice@bbsrc.ac.uk
COMMENT   Vector pBLUESCRIPT SK+
            R. Site 1 Ecorl
            R. Site 2 Ecorl
            Contact: David Morrice
            Dept. Genomics and Bioinformatics
            Roslin Institute, Roslin
            Midlothian, EH25 9PS, UK
            Tel: +44 (0) 131 527 4200
            Fax: +44 (0) 131 440 0434
            Email: david.morrice@bbsrc.ac.uk
            Seq primer T3.
FEATURES   Location/Qualifiers
   source   1..680
            /organism="Meleagris gallopavo"
            /db_xref="taxon:9103"
   STS   1..680
            /standard_name="TC1078"
BASE COUNT   195 a   131 c   163 g   191 t
ORIGIN

```
Db    667 TTGACCTCT 675
```

Query Match   14.38; Score 71.4; DB 11; Length 680;
Best Local Similarity 72.1%; Pred. No. 1.7e-08;
Matches 93; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy    363 ATCAAAGCAGCTCATAAAAGCAATGGTTTTAAAACATCACCCAGACAAAACGGAAAGCAGCT 422
          ||| |      |      || |||||| | ||| |||||||| ||| ||| | |||||| ||
Db    547 ATCTACTTGCCAGATAAAATCCAATGGTTCTGAAACATCATCCAGACAAGGCGAAAAGCTGCA 606

Qy    423 GGTGAACCAAATAAAAGAGGAGAATAAATGACTACTTCACTTGCATAACTAAAGCTTATGAA 482
          |||| |      ||||||||||||||||||||   |||||||||||||| | ||||| |||
Db    607 GGGGAGCAAAATAGGAGAGAAGGTGATAAATGATTATTTTACGTGCATAACTAAAGGTAAGTCA 666

Qy    483 ATGTTATCT 491
          || |||
Db    607 ... 
```

RESULT 15
LOCUS   AL593453/c
AL593453/c
DEFINITION Meleagris gallopavo DNA; STS TC1078, sequence tagged site.
ACCESSION   AL593453
VERSION   AL593453.1   GI:14716571
KEYWORDS   STS.
SOURCE   turkey.
ORGANISM   Meleagris gallopavo
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Meleagrididae;
            Meleagris.
REFERENCE   1 (bases 1 to 680)
AUTHORS   Morrice,D.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 680)
AUTHORS   Morrice,D.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Morrice D., Roslin Institute, Roslin,
            Midlothian, EH25 9PS, UK david.morrice@bbsrc.ac.uk
COMMENT   Vector pBLUESCRIPT SK+
            R. Site 1 Ecorl
            R. Site 2 Ecorl
            Contact: David Morrice
            Dept. Genomics and Bioinformatics
            Roslin Institute, Roslin
            Midlothian, EH25 9PS, UK
            Tel: +44 (0) 131 527 4200
            Fax: +44 (0) 131 440 0434
            Email: david.morrice@bbsrc.ac.uk
            Seq primer T7.
FEATURES   Location/Qualifiers
   source   1..680
            /organism="Meleagris gallopavo"
            /db_xref="taxon:9103"
   STS   1..680
            /standard_name="TC1078"
BASE COUNT   190 a   163 c   195 t   1 others
ORIGIN

Query Match   14.38; Score 71.4; DB 11; Length 680;
Best Local Similarity 72.1%; Pred. No. 1.7e-08;
Matches 93; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy    363 ATCAAAGCAGCTCATAAAAGCAATGGTTTTAAAACATCACCCAGACAAAACGGAAAGCAGCT 422
          ||| |      |      || |||||| | ||| |||||||| ||| ||| | |||||| ||
Db    134 ATCTACTTGCCAGATAAAATCCAATGGTTCTGAAACATCATCCAGACAAGGCGAAAAGCTGCA 75

Qy    423 GGTGAACCAAATAAAAGAGGAGAATAAATGACTACTTCACTTGCATAACTAAAGCTTATGAA 482
          |||| |      ||||||||||||||||||||   |||||||||||||| | ||||| |||
Db     74 GGGGAGCAAAATAGGAGAGAAGGTGATAAATGATTATTTTACGTGCATAACTAAAGGTAAGTCA 15

Qy    483 ATGTTATCT 491
          || |||
Db     14 TTGACCTCT 6
```

Search completed: October 10, 2002, 17:46:46
Job time : 1046.87 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
(without alignments)
14616.225 Million cell updates/sec

Title: us-09-489-101A-9_COPY_1_500
Perfect score: 500
Sequence: 1 gaattcctgactgccacagg.........cgatatgattactggatg 500
Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 1046326293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*

```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*
```

SUMMARIES

Result No. | Score | Query Match | Length | DB | ID | Description

No.		Score	Query Match	Length	DB	ID	Description
1		500	100.0	2291	9	AX201588	AX201588 Sequence
2		500	100.0	2291	9	HSU23028	U23028 Human eukar
3		500	100.0	2651	9	BC013590	BC013590 Homo sapi
4	c	405.6	81.1	2598	4	OCU23037	U23037 Oryctolagus
5		376.8	75.4	2488	10	RNU19516	U19516 Rattus norv
6	c	262.2	52.4	151032	2	AC061705	AC061705 Homo sapi
7	c	262.2	52.4	152623	2	AC024530	AC024530 Homo sapi
8	c	262.2	52.4	229586	2	AC048331	AC048331 Homo sapi
9	c	251.4	50.3	298408	2	AC092931	AC092931 Homo sapi
10	c	251.4	50.3	199359	2	AC074027	AC074027 Mus muscu
11	c	182.8	36.6	239574	2	AC073791	AC073791 Mus muscu
12	c	180.6	36.1	13164	10	RNU19511	U19511 Rattus norv
13		81.4	16.3	2542	9	AK056400	AK056400 Homo sapi
14	c	71.8	14.4	199359	2	AC074027	AC074027 Mus muscu
15		70.2	14.0	2903	8	YSCGCDA	YSCGCDA Saccharomyc
16		70.2	14.0	20089	8	SC81428	SC81428 S.cerevisia
17		70.2	14.0	39827	8	SC8142B	SC8142B S.cerevisia
18		65.4	13.1	36310	8	Z68194	Z68194 S.cerevisia
19	c	38.8	7.8	2007	3	Z99168	Z99168 S.pombe chr
20		38.8	7.8	29352	3	AF289103	AF289103 Drosophil
21	c	38.8	7.8	48287	3	DMC86E4	DMC86E4 Drosophil
22	c	38.8	7.8	169618	3	AL021086	AL021086 Drosophil
23	c	38.8	7.8	170490	3	AC017774	AC017774 Drosophil
24	c	38.8	7.8	300739	3	AC105055	AC105055 Drosophil
25	c	37.2	7.4	162112	2	AE003422	AE003422 Drosophil
26	c	37.2	7.4	184810	2	AC104147	AC104147 Drosophil
27	c	36.4	7.3	164968	2	AC011572	AC011572 Drosophil
28	c	36.4	7.3	170669	2	AC096732	AC096732 Homo sapi
29	c	36.4	7.3	265861	2	AC040164	AC040164 Homo sapi
30	c	35.2	7.0	125020	2	AC007907	AC007907 Homo sapi
31	c	35.2	7.0	198740	2	AC069274	AC069274 Mus muscu
32	c	35	7.0	43320	2	AF429315	AF429315 Homo sapi
33	c	35	7.0	148656	2	AC069434	AC069434 Homo sapi
34		35	7.0	224297	2	AL513215	AL513215 Human DNA
35	c	34.8	7.0	87400	2	AC005482	AC005482 Homo sapi
36		34.8	7.0	170518	2	AC027012	AC027012 Homo sapi
37		34.6	6.9	169029	2	AC010796	AC010796 Homo sapi
38		34.6	6.9	200109	2	AC091787	AC091787 Oryza sat
39		34.4	6.9	37098	3	CNS0C7R	CNS0C7R Human chr
40		34.4	6.9	139788	2	CNS07EE2	CNS07EE2 Human chr
41		34.4	6.9	159305	2	U28737	U28737 Caenorhabdi
42	c	34.4	6.9	193483	2	AC024118	AC024118 Homo sapi
43	c	34.4	6.9	210791	2	AL355384	AL355384 Human DNA
44		34.4	6.9	210791	2	AL356257	AL356257 Human DNA
45	c	34.2	6.8	110127	9	HS789011	AL031075 Human DNA

RESULT 1
AX201588

```
LOCUS       AX201588                2291 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 9 from Patent WO0153349.
ACCESSION   AX201588
VERSION     AX201588.1  GI:15391437
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2291)
AUTHORS     Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
TITLE       Small cell lung cancer associated antigens and uses therefor
JOURNAL     Patent: WO 0153349-A 9 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..2291
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

ALIGNMENTS

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
FEATURES
    source          1..2291
                    /organism="Homo sapiens"
                    /note="U-937 histiocytic lymphoma"
                    /cell_line="U-937 histiocytic lymphoma"
```

```
BASE COUNT     588 a     540 c     630 g     533 t
ORIGIN

Query Match      100.0%;    Score 500;  DB 6;  Length 2291;
Best Local Similarity   100.0%;    Pred. No. 5.6e-143;
Matches 500;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 GAATTCCTGACTCCACAGGTGTACACAGGAAAACATTTGTTCTTTGTTGTTGTTGCTTGGAAAGCTGCT   60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GAATTCCTGACTCCACAGGTGTACACAGGAAAACATTTGTTCTTTGTTGTTGTTGCTTGGAAAGCTGCT   60

Qy   61 CAAATCAAAAGAACACATTTACTGAAGTCAAAGTGGTGGTGCCGCCCTACACATCTCAATGTGGTT  120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAAATCAAAAGAACACATTTACTGAAGTCAAAGTGGTGGTGCCGCCCTACACATCTCAATGTGGTT  120

Qy  121 CGAATAATTACATCAGAGCTCTATCGATCACACTGGGAGATGTCCTCCGTGATGTTGATGCC      180
Db  121 CGAATAATTACATCAGAGCTCTATCGATCACACTGGGAGATGTCCTCCGTGATGTTGATGCC      180

Qy  181 AAGGCTTTGGTCGCGCTCTGACTTTCTTCTGGGTGTATGGGGATGGGATGTCATCTCAAAACATCAAT 240
Db  181 AAGGCTTTGGTCGCGCTCTGACTTTCTTCTGGGTGTATGGGGATGGGATGTCATCTCAAAACATCAAT 240

Qy  241 ATCACCAGAGACGCCCTTGAGGACACACAGGTTGAGGACGGAAGCTAGAAAAAAATGTTTCTGTG 300
Db  241 ATCACCAGAGACGCCCTTGAGGACACACAGGTTGAGGACGGAAGCTAGAAAAAAATGTTTCTGTG 300

Qy  301 ATGACGACGATCTTCAAGGAGAGTCATCCCCAGCCACCCAACTCGTTGCCACGAAGACAAT  360
Db  301 ATGACGACGATCTTCAAGGAGAGTCATCCCCAGCCACCCAACTCGTTGCCACGAAGACAAT  360

Qy  361 GTGGTAGTGGGCTGTGGATAGTACCACAAAACAGGGTTCTCCATTTTCAGAAGACCCAGGGT  420
Db  361 GTGGTAGTGGGCTGTGGATAGTACCACAAAACAGGGTTCTCCATTTTCAGAAGACCCAGGGT  420

Qy  421 CTCCGGCGTTTTGCATTTCCTCTGAGCCTGTTTCAGGGCCAGTAGTGATGGAGTGGAGGTT 480
Db  421 CTCCGGCGTTTTGCATTTCCTCTGAGCCTGTTTCAGGGCCAGTAGTGATGGAGTGGAGGTT 480

Qy  481 CGAATATGATTTACTGGATTG 500
Db  481 CGAATATGATTTACTGGATTG 500
```

```
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2651)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (04-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT     NIH-MGC Project URL: http://mgc.nci.nih.gov
REMARK      Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 14 Row: a Column: 2.
FEATURES             Location/Qualifiers
     source          1..2651
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:9947 IMAGE:3876105"
                     /tissue_type="Lung, large cell carcinoma"
                     /clone_lib="NIH_MGC_68"
                     /lab_host="DH10B"
                     /vector="pCMV-SPORT6"
     CDS             51..2216
                     /codon_start=1
                     /product="Unknown (protein for MGC:9947)"
                     /protein_id="AAH13590.1"
                     /db_xref="GI:15488926"
                     /translation="MAAPVVAPPGVVVSRANKRSGAGPGGSGGGGARGAEEEPPPLQ
                     AVLVADSFDRFFPISKDQPRVLLPLANVALIDYTLEFLIATGVQETFVFCCWKAAQI
                     TRALEEHRLRRKLEKNVSVWTMIFKESSPSHPTRCHEDNVVAVDSTTNRVLHFQKT
                     KEHLLKSWKCRPSLNVVRIITSELYRSLGDVLRDVDAKALVRSDFLLVGVDVISNIN
                     QGLRRFAFPLSLFQGSSDGVEVKRVTLKPRSVLITSQVVVGPNITLFPGSVIS
                     LVNEEILGNQIHHVTAKEYGARVSNLHMYSAVCADVIRRWVPLTPEANFTDSTTQS
                     CTHSRHNIYRGPEVSLGHGSILEENVLLGGGTVIGSNCFITNSVIGPGCHIGDNVVLD
                     QYILWQGVRVAAGAQIHQSLLCDNAEVKRVTLKPRSVLITSQVVVGPNITLFPGSVIS
                     LHPPDAEEDEDGEFSDDSGADQEKDKVKMKGYNPAEVGAAGKGYLWKAAGMNMEEE
                     ELQQNLWGLKINMEEESESSEOSMDSEEPDSRGGSPOMDIKVFQNEVLGTLQRGKE
                     ENTSCDNLVLEINSLKYAYNVSLKEVNVQVLSHVLEFPLQOMDSPLDSSRYCAILLPL
                     LKAWSPVFRNYIKRAADHLEALAAIEDFFLEHEALGISMAKVLMAFYQLELIAEETIL
                     SWFSQRDTTDKGQOLRKNQOLQRFIQWLKEAEEESSEDD"
```

```
RESULT 4

OCU23037
LOCUS       OCU23037     2598 bp    mRNA    linear    MAM 22-AUG-1996
DEFINITION  Oryctolagus cuniculus eukaryotic initiation factor 2B-epsilon mRNA,
            complete cds.
ACCESSION   U23037
VERSION     U23037.1  GI:806855
KEYWORDS    .
SOURCE      domestic rabbit.
ORGANISM    Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 2598)
AUTHORS     Asuru,A.I., Mellor,H., Thomas,N.S., Yu,L., Chen,J.J., Crosby,J.S.,
            Hartson,S.D., Kimball,S.R., Jefferson,L.S. and Matts,R.L.
TITLE       Cloning and characterization of cDNAs encoding the epsilon-subunit
            of eukaryotic initiation factor-2B from rabbit and human
JOURNAL     Biochim. Biophys. Acta 1307 (3), 309-317 (1996)
MEDLINE     96305354
REFERENCE   2  (bases 1 to 2598)
AUTHORS     Asuru,A.I., Mellor,H., Thomas,N.S.B., Yu,L., Chen,J.-J.,
            Crosby,J.S., Hartson,S.D., Kimball,S.R., Jefferson,L.S. and
            Matts,R.L.
TITLE       Direct Submission
JOURNAL     Submitted (17-MAR-1995) Robert L. Matts, Biochemistry & Molecular
            Biology, Oklahoma State University, 246 NRC, Stillwater, OK
            74078-0454, USA
FEATURES             Location/Qualifiers
     source          1..2598
                     /organism="Oryctolagus cuniculus"
                     /strain="New Zealand White"
                     /db_xref="taxon:9986"
                     /cell_type="reticulocyte"
     CDS             72..2237
                     /codon_start=1
                     /note="epsilon subunit of eukaryotic initiation factor 2B"
                     /product="eIF-2Bepsilon"
                     /protein_id="AAC48618.1"
                     /db_xref="GI:806856"
                     /translation="MATTVVAPPGAVSDRANKRGGGPGGGGGGGGGARGAEEESPPLQ
                     AVLVADSFNRRFPISKDQPRVLLPLANVALIDYTLEFLIATGVQETFVFCCWKAAQI
                     KEHLQKSWKCRPTSLNVVRIITSELYRSLGDVLRDVDAKALVRSDFLLVGDVVSNIN
                     VTRALEEHRLRRKLEKNVSVMIMIFKESSPSHPTRCHEDNVVAVDSSATNRILHFDY
                     OGLRRFSFPLSLFQGSSGAGVEIRYDLLDCHISICSPQVAQLFTDNFDYQTRDDFVRGL
                     LVNEEILGNOIHHVHVTREYGARVSNLHHYSAVCADVIRRWVFPLTPEANFTDSTAQS
                     CTHSRHNIYRGPEVSLGHGSILEENVLLGSGTVIGSNCSITNSVIGPGCCIGDNVVLD
                     RAYLWKGVQVASAGAQIHQSLLCDHAEVKEQVTLKPHCVLTSQVVVGNITLPEGSVIS
                     LHPPDAEEDEDGQFSDDSGSVGNOAKERAKLKGYNPAEVGVAGKGYLWKAADMNTEKEE"
```

```
BASE COUNT      705 a     623 c     737 g     586 t
ORIGIN

Query Match      100.0%;      Score 500;      DB 9;      Length 2651;
Best Local Similarity    100.0%;
Matches  500;      Conservative   0;      Mismatches   0;      Indels   0;      Gaps   0;      Pred. No. 5.7e-143;

Qy      1 GAATTCCTGACTGCCACAGGTGTACAGGAGGAAACATTGTTCTTTTGTTCTTTTGCTGGAAAGCTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 GAATTCCTGACTGCCACAGGTGTCACAGGAGAAACATTTGTTCTTTTGTTCTTTTGCTGGAAAGCTGCT 350

Qy     61 CAAATCAAAGAACACATTTACTGAAGTGAAGTGGTGCGCGCCCCTACACTCTCAATGTCGGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 CAAATCAAAGAACACATTTACTGAAGTGAAGTGGTGCGCGCCCCTACACTCTCAATGTCGGTT 410

Qy    121 CGAATAAATTACATCAGAGCTCTATCGATCACTGGGAGAGATGTCCTCCGTGATGTTGATGCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 CGAATAAATTACATCAGAGCTCTATCGATCACTGGGAGAGATGTCCTCCGTGATGTTGATGCC 470

Qy    181 AAGGCTTTTGGTGCGCCTCTGAACTTTTCTTCTTGGTGTTGATGGGGAATGTGATGGGGAAGATGTCARCTCAAACATCAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 AAGGCTTTTGGTGTCGCGCTCTGACTTTTCTTCTTCGGTGTTGATGGGGAATGTCATCTCAAACATCAAT 530

Qy    241 ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAGCTAGAAAAAATGTTTCTGTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAGCTAGAAAAAATGTTTCTGTG 590

Qy    301 ATGACGATGATCTTCAAGGAGTCATCCCCCAGCCACCCGACCTCGTTGCCACCGAAGACAAT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 ATGACGATGATCTTCAAGGAGTCATCCCCCAGCCACCCGACCTCGTTGCCACCGAAGACAAT 650

Qy    361 GTGTAGTGGCTGGCTGGGATAGTACCACACAAAACAGGGTTCTCCATTTTCAGAGAGACCCAGGT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 GTGGTAGTGGCTGGCTGGGATAGTACCACACAAAACAGGGTTCTCCATTTTCAGAGAGACCCAGGT 710

Qy    421 CTCCCGGCGTTTTGCATTCCCTGAGCCTGTTTCACGGCACGTAGTGATTGGGAGTGGGAGGTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 CTCCCGGCGTTTTGCATTTCCTCTGCGACCCTGTTTCACGGCACGTAGTGATTGGGAGTGGGAGGTT 770

Qy    481 CGAATATGATTTACTGGATTG 500
          |||||||||||||||||||||
Db    771 CGAATATGATTTACTGGATTG 790
```

BASE COUNT   606 a   675 c   761 g   556 t
ORIGIN

```
Qy        1 GAATTCCTGACTGCCACAGGTGTACAGGAGAAACATTTGTTTTTTGTTGCTGGAAAGCTGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      312 GAATTCCTGACTGCCACAGGTGTACAGGAGAAACATTTGTTTTTTGTTGCTGGAAAGCTGCT 371

Qy       61 CAAATCAAAGAACATTTACTGAAGTCAAAGTGGTGGCGCCCTTACATTCTCAATGTGGTT 120
            |||||||||||||||||||||||||||||||||| | ||||||||||||||||||||||||
Db      372 CAGATCAAAGAACATTTGCAAAAATCCAAGTGGTGCCGCCCCAACATCCCTCAACGTGGTT 431

Qy      121 CGAATAAATTACATCAGAGCTGTCTATGGAGATGTCCTCCGTGATGTTGATGCC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      432 CGGATAAATTACATCAGAGAGTCTTACCGGATCCGCTGGGAGATGTCCTCCGTGATGTTGACGCC 491

Qy      181 AAGGCTTTGGTGCGCTCTGACTTTCTTCTTCGGTGTATGGGATGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||
Db      492 AAAGCCCTGTGTGCGCTCTGACTTCCTCTGTGGTGTACGGAGATGTCGTCTCCAACATCAAT 551

Qy      241 ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAAGCTAGAAAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      552 GTCACCAGAGCCCTCGAGGAACACAGAGGTTAAGCGGGAAGCTAGAAAAAAAATGTTCTGTG 611

Qy      301 ATGACGATGATCTTCAAGGAGTCATCCCCAGCCACCCACACTCGTTGCCACGAAGACAAT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      612 ATGACAATGATCTTCAAAGAATCATCCCCCAGCCGCTTGCCATGAGGACAAT 671

Qy      361 GTGGTAGTGCGGCTGGGATAGTACACAACACAGGGTTCTCCATTTTCAGAAGACCCCAGGT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      672 GTGGTAGTGCGGCTGGGATAGTGCCACAAACCGGATTCTCCACTTCCAGAAGACCCAGGC 731

Qy      421 CTCCGGCCGTTTGCATTTCCTCTGAGCCTGTTCAGGGAGTGATGGAGTGGAGGTT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      732 CTCCGACGTTTCTCTTTCCTCTGAGCCTGTGGCCGCGTGGGCCGTGGGGAGATT 791

Qy      481 CGAATATGATTTACTGGATTG 500
            |||||||||||||||||||||
Db      792 CGGTATGACTTATTGGATTG 811
```

RESULT 5
LOCUS   RNU19516   2488 bp   mRNA   linear   ROD 30-OCT-1996
DEFINITION   Rattus norvegicus initiation factor eIF-2Be mRNA, complete cds.
ACCESSION   U19516
VERSION   U19516.1   GI:924598
KEYWORDS   .
SOURCE   Norway rat.
  ORGANISM   Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE   1 (bases 1 to 2488)
  AUTHORS   Flowers,K.M., Mellor,H., Matts,R.L., Kimball,S.R. and
           Jefferson,L.S.
  TITLE   Cloning and characterization of complementary and genomic DNAs
           encoding the epsilon-subunit of rat translation initiation
           factor-2B
  JOURNAL   Biochim. Biophys. Acta 1307 (3), 318-324 (1996)
  MEDLINE   96305355
REFERENCE   2 (bases 1 to 2488)
  AUTHORS   Mellor,H.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-JAN-1995) Harry Mellor, Cellular Molecular
           Physiology, Penn State College of Medicine, Milton S. Hershey

Medical Center, Hershey, PA 17033, USA
FEATURES   Location/Qualifiers
     source   1..2488
                 /organism="Rattus norvegicus"
                 /strain="Sprague Dawley"
                 /db_xref="taxon:10116"
                 /clone_lib="rat brain cDNA in UNI-ZAP XR (stratagene)"
     CDS   35..2185
                 /function="initiation factor"
                 /codon_start=1
                 /product="eIF-2Be"
                 /protein_id="AAB17690.1"
                 /db_xref="GI:924599"
                 /translation="MAATAAVPSAVGGRANKRGGGSGGGGTTGGAEEPPPLQAVLVA
                 DSFDRRFFPISKDQPRVLLPLANVALIDYTLEFLTAYGVQEFTFVFCCNKAAQIKEHLQ
                 KSKWCHPTSLNVVRITTTSDIYRSLGDVLRDVDAKALVRSDFLIYGDVVSNINISKAL
                 EEHRLRRKLEKNVSVMWFKESSSPHPTRCHEDNVVLAVDSTTNRILHFQKTQGLRH
                 FSFPLGFEQGSLDGVEIRYDLLDCHISICSPQVAQLFTDNFDYQTRDDFVRGLLVNEE
                 ILGNQIHLHVTSREYGSRVSNLHMYSAVCCDVIRRNWYTLPEVPNEFHFDTSSTOSYTHSR
                 HNIYRGPEVSLGHGSVLEENVLLGAGTVVGSNCSIPNSVIGPNCHIGDNVVLDQAYLW
                 QGVRVAAGRQIHQSLLCDRAEVKERVILPHGVLTSQVVVGPDIILPEGSVISLHPPD
                 AEEDEDDGQFSDDSGADQOREKRVKLKGYNPAEVGPEGQGYLWKAEVEDKEDEELRQS
                 LWGLMIMMEEESETESERSRVDPEELLDSRAGSPOLDDIRVPQNEVLGTLQRGREENISC
                 DNLVLEINSLKYAVNISLKEVMQVLSHVVLEPPLQOVDGVLDPNRYCALLLPLLKAWS
                 PVFRNYIKRAADHLEALAAIEEFFLEHEALGTICIAKVLMGFYQLEILAEETILLSWFSQ
                 RDITDKGQQLRKNQOLQRFTQWLRAEEEESSDDD"
     polyA_signal   2449..2454
     polyA_site   2488
                 /note="30 A nucleotides"

BASE COUNT   628 a   630 c   678 g   552 t
ORIGIN

Query Match   75.4%;   Score 376.8;   DB 10;   Length 2488;
Best Local Similarity   84.6%;
Matches 423;   Conservative   0;   Mismatches   77;   Indels   0;   Gaps   0;

```
Qy        1 GAATTCCTGACTGCCACAGGTGTACAGGAGAAACATTTGTTTTTTGTTGCTGGAAAGCTGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      260 GAATTCCTGACTGCCACGGTGTACAGGAGAAACATTTGTTTTTTGTTGCTGGAAAGCTGCT 319

Qy       61 CAAATCAAAGAACATTTACTGAAGTCAAAGTGGTGGCGCCCTTACATTCTCAATGTGGTT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      320 CAGATCAAAGAACAACTTACAGAAAATCCAAATGGTCGCCCTACATTCTCTCAATGTGGTTT 379

Qy      121 CGAATAAATTACATCAGAGCTGTCTATCGGAGATGTCCTCCGTGATGTTGATGCC 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      380 CGGATAAATTACATCGGAGAGTCTTACCGGATCACTGGGAGATGTCCTCCGTGATGTTGTT 439

Qy      181 AAGGCTTTGGTGCGCTCTGACTTTCTTCTTCGGTGTATGGGATGAGTCATCTCAAAACATCAAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      440 AAGGCCCTTGGTGCGCTCTGACTTCCTCCTCGATATAGGGATGTCGTCTCCGAATATACAAT 499

Qy      241 ATCACCAGAGCCCTTGAGGAACACAGGTTGAGACGGAAAGCTAGAAAACACAGGTT 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      500 ATTCCAAAGCCCTGGAGGAGAGACACAGGTTAAGGAAGCTAGAAAAAAAGTGTTCTGTC 559

Qy      301 ATGACGACGATCTTCAAGGAGTCATCCCCAGCCACCCACACTCGTTGCCACGAAGACAAT 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      560 ATGACAAATGTGTTCAAAGAGTCCACCACGGTCCCATGAGGAGAAGACAAT 619

Qy      361 GTGGTAGTGCGGCTGGGATAGTACCACAAACACAGGGTTCTCCATTTTCAGAAGACCCAGGT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 GTGGTGCTAGCGTGGATAGCACACAAAACAGGAATTCTTCACTTCACAGAAGACCCAGGC 679

Qy      421 CTCCGGCCGTTTGCATTTCCTCTGAGCCTGTTCAGGGCAGCGCAGTAGTGATGGGAGTGGAGGTT 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 CTCCGCCATTCTCGTTCCTTTGGGCCTATTCCAGGGCCAGTCTAGACGGAGCAGTGGGAGATT 739

Qy      481 CGAATATGATTTACTGGATTG 500
            |||||||||| || ||||||
Db      740 CGAATATGACTTGTGGGATTG 759
```

```
RESULT 6
LOCUS       AC061705/c
LOCUS       AC061705   151032 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-488M12, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC061705
VERSION     AC061705.14  GI:18449515
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151032)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,E., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Jackson,L.E., Jacobson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Karlsson,E., Mapua,P., Martindale,A., Martinez,E.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151032)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 31, 2002 this sequence version replaced gi:9743372.
            ----------------------------------------------------------
            -Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ----------------------------------------------------------
            -Project Information
              Center project name: HANK
              Center clone name: RP11-488M12
            ----------------------------------------------------------
            -Summary Statistics
              Sequencing vector: M13; L08821
              Chemistry: Dye-primer BoIdpy: 18% of reads
              Chemistry: Dye-terminator Big Dye: 82% of reads
              Assembly program: Phrap; version 0.990329First call to
            findPhraplist
              Consensus quality: 140075 bases at least Q40
              Consensus quality: 148512 bases at least Q30
              Consensus quality: 153337 bases at least Q20
              Estimated insert size: 149429; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
            *
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1    44630: contig of 44630 bp in length
            *    44631    44730: gap of unknown length
            *    44731    65344: contig of 20614 bp in length
            *    65345    65444: gap of unknown length
            *    65445    75360: contig of 9916 bp in length
            *    75361    75460: gap of unknown length
            *    75461    84626: contig of 9166 bp in length
            *    84627    84726: gap of unknown length
            *    84727    93423: contig of 8697 bp in length
            *    93424    93523: gap of unknown length
            *    93524   100831: contig of 7308 bp in length
            *   100832   100931: gap of unknown length
            *   100932   107761: contig of 6830 bp in length
            *   107762   107861: gap of unknown length
            *   107862   115279: contig of 7418 bp in length
            *   115280   115379: gap of unknown length
            *   115380   120595: contig of 5216 bp in length
            *   120596   120695: gap of unknown length
            *   120696   128475: contig of 7780 bp in length
            *   128476   128575: gap of unknown length
            *   128576   134874: contig of 6299 bp in length
            *   134875   134974: gap of unknown length
            *   134975   140581: contig of 5607 bp in length
            *   140582   140681: gap of unknown length
            *   140682   143980: contig of 3299 bp in length
            *   143981   144080: gap of unknown length
            *   144081   146393: contig of 2313 bp in length
            *   146394   146493: gap of unknown length
            *   146494   148710: contig of 2217 bp in length
            *   148711   148810: gap of unknown length
            *   148811   151032: contig of 2222 bp in length
FEATURES             Location/Qualifiers
     source          1..151032
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-488M12"
                     /chromosome="3"
BASE COUNT    37567 a  38444 c  38131 g  35347 t   1543 others
ORIGIN
```

```
Query Match      52.4%;       Score 262.2;  DB 2;   Length 151032;
Best Local Similarity    79.4%;   Pred. No. 1.8e-69;
Matches 366;     Conservative    0;   Mismatches  3;   Indels  92;  Gaps  1;

Qy         76 TTACTGAAGTCAAAAGTGGTGCCGCCCCTACACTCTCTCAAATGTGGTTCGGAATAATTACACTCA 135
              |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||||||
Db     102958 TTTAGGAAGTCAAAAGTGGTGCCGCCCCTACACTCTCTCAATGTGGTTCGGAATAATTACATCA 102899

Qy        136 GAGCTCTATCGATCACTGGGAGAGATGTCCTCCCGTGATTGTTGATGATCCCAAGGCTTTGGTGCGC 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102898 GAGCTCTATCGATCACTGGGAGATGTCCTCCCGTGATTGTTGATGCCAAGGCTTTGGTGCGC 102839

Qy        196 TCTGACTTTCTTCTGGTGTATGGGGATGTCATCTCAAACATCAATATCACCAGAGCCCTT 255
```

```
Db 102838  TCTGACTTCTTCTGGTGTATGGGAGTCATCTCAAACATCATAATATCACCAGAGCCCTT 102779
           ||||||||||||||||||||
Qy    256  GAGGAACAC----------                                          264
           |||||||||
Db 102778  GAGGAACACAGGTCAGGATGGGAAAATGACAGGAACACAAGGGTTAAAGACCAGCAGAGCC 102719
           |||||||||||||||||||||||||||||||||||||||||
Qy    265  -----------------------------------AGGTTGAGACGGAAGCTAG         283
                                              |||||||||||||||||
Db 102718  TGAGACTGCTTTTGTGCAGTTGTGTCCTCCTGTCCTTTATAGGTTGAGACGGAAGCTAG   102659
           ||||||||||||||||||
Qy    284  AAAAAAATGTTTCTGTGATGACGATGATCTTCAAGGAGTCAT                   343
           ||||||||||||||||||||||||||||||||||||||||||
Db 102658  AAAAAAATGTTTCTGTGATGACGATGATCTTCAAGGAGTCATCCCCAGCCCACCCAACTC 102599

Qy    344  GTTGCCACGAAGACAAATGTGGTAGTGGCTGTGGATAGTAGTCACACAAACAGGGGTTCTTCCATT 403
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 102598  GTTGCCACGAAGACAAATGTGGTAGTGGCTGTGGATAGTAGTACCACACAAACAGGGGTTCTTCCATT 102539
Qy    404  TTCAGAAGACCCAGGGTCTCCGGCGTTTTTGCATTTCCTCTG                   444
           ||||||||||||||||||||||||||||||||||||||||||
Db 102538  TTCAGAAGACCCAGGGTCTCCGGCGTTTTTGCATTTCCTCTG 102498
```

```
RESULT 7
LOCUS       AC024530/c           152623 bp    DNA    linear    HTG 14-MAR-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-131020 map 4, *** SEQUENCING
            IN PROGRESS ***, 49 unordered pieces.
ACCESSION   AC024530
VERSION     AC024530.3  GI:7239669
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152623)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 4, clone RP11-131020
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 152623)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,M., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campoplano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N, Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illey,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P, O'Neil,D, Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Triglilo,J., Vassilev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 14, 2000 this sequence version replaced gi:7229935.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
```

```
            Contact: sequence_submissions@genome.wi.mit.edu
            --------- Project Information
            Center project name: L7361
            Center clone name: 131_O_20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*        1  1156: contig of 1156 bp in length
*     1157  1256: gap of 100 bp
*     1257  2440: contig of 1184 bp in length
*     2441  2540: gap of 100 bp
*     2541  3807: contig of 1267 bp in length
*     3808  3907: gap of 100 bp
*     3908  5016: contig of 1109 bp in length
*     5017  5116: gap of 100 bp
*     5117  6229: contig of 1113 bp in length
*     6230  6329: gap of 100 bp
*     6330  7466: contig of 1137 bp in length
*     7467  7566: gap of 100 bp
*     7567  8835: contig of 1269 bp in length
*     8836  8935: gap of 100 bp
*     8936 10247: contig of 1312 bp in length
*    10248 10347: gap of 100 bp
*    10348 11428: contig of 1081 bp in length
*    11429 11528: gap of 100 bp
*    11529 12880: contig of 1352 bp in length
*    12881 12980: gap of 100 bp
*    12981 13991: contig of 1011 bp in length
*    13992 14091: gap of 100 bp
*    14092 15759: contig of 1668 bp in length
*    15760 15859: gap of 100 bp
*    15860 17156: contig of 1297 bp in length
*    17157 17256: gap of 100 bp
*    17257 18728: contig of 1472 bp in length
*    18729 18828: gap of 100 bp
*    18829 20169: contig of 1341 bp in length
*    20170 20269: gap of 100 bp
*    20270 21411: contig of 1142 bp in length
*    21412 21511: gap of 100 bp
*    21512 22964: contig of 1453 bp in length
*    22965 23064: gap of 100 bp
*    23065 25024: contig of 1960 bp in length
*    25025 25124: gap of 100 bp
*    25125 27388: contig of 2264 bp in length
*    27389 27488: gap of 100 bp
*    27489 28881: contig of 1393 bp in length
*    28882 28981: gap of 100 bp
*    28982 30874: contig of 1893 bp in length
*    30875 30974: gap of 100 bp
*    30975 33014: contig of 2040 bp in length
*    33015 33114: gap of 100 bp
*    33115 34935: contig of 1821 bp in length
*    34936 35035: gap of 100 bp
*    35036 37222: contig of 2187 bp in length
*    37223 37322: gap of 100 bp
*    37323 39483: contig of 2161 bp in length
*    39484 39583: gap of 100 bp
*    39584 41860: contig of 2277 bp in length
*    41861 41960: gap of 100 bp
*    41961 44448: contig of 2488 bp in length
*    44449 44548: gap of 100 bp
*    44549 46452: contig of 1904 bp in length
*    46453 46552: gap of 100 bp
*    46553 49889: contig of 3337 bp in length
*    49890 49989: gap of 100 bp
*    49990 52994: contig of 3005 bp in length
*    52995 53094: gap of 100 bp
```

```
FEATURES             Location/Qualifiers
     source          1..152623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-131020"
     misc_feature    1..1156
                     /note="assembly_fragment"
     misc_feature    1257..2440
                     /note="assembly_fragment"
     misc_feature    2541..3807
                     /note="assembly_fragment"
     misc_feature    3908..5016
                     /note="assembly_fragment"
     misc_feature    5117..6229
                     /note="assembly_fragment"
     misc_feature    6330..7466
                     /note="assembly_fragment"
     misc_feature    7567..8835
                     /note="assembly_fragment"
     misc_feature    8936..10247
                     /note="assembly_fragment"
     misc_feature    10348..11428
                     /note="assembly_fragment"
     misc_feature    11529..12880
                     /note="assembly_fragment"
     misc_feature    12981..13991
                     /note="assembly_fragment"
     misc_feature    14092..15759
                     /note="assembly_fragment"
     misc_feature    15860..17156
                     /note="assembly_fragment"
     misc_feature    17257..18728
                     /note="assembly_fragment"
     misc_feature    18829..20169
                     /note="assembly_fragment"
     misc_feature    20270..21411
                     /note="assembly_fragment"
     misc_feature    21512..22964
                     /note="assembly_fragment"
     misc_feature    23065..25024
                     /note="assembly_fragment"
```

```
* 53095  55588: contig of 3494 bp in length
* 56589  56688: gap of 100 bp
* 56689  59956: contig of 3268 bp in length
* 59957  60056: gap of 100 bp
* 60057  63904: contig of 3848 bp in length
* 63905  64004: gap of 100 bp
* 64005  67438: contig of 3434 bp in length
* 67439  67538: gap of 100 bp
* 67539  70684: contig of 3146 bp in length
* 70685  70784: gap of 100 bp
* 70785  75479: contig of 4695 bp in length
* 75480  75579: gap of 100 bp
* 75580  79243: contig of 3664 bp in length
* 79244  79343: gap of 100 bp
* 79344  82983: contig of 3640 bp in length
* 82984  83083: gap of 100 bp
* 83084  87910: contig of 4827 bp in length
* 87911  88010: gap of 100 bp
* 88011  94315: contig of 6305 bp in length
* 94316  94415: gap of 100 bp
* 94416  99313: contig of 4898 bp in length
* 99314  99413: gap of 100 bp
* 99414  104977: contig of 5564 bp in length
* 104978 105077: gap of 100 bp
* 105078 112571: contig of 7494 bp in length
* 112572 112671: gap of 100 bp
* 112672 118896: contig of 6225 bp in length
* 118897 118996: gap of 100 bp
* 118997 125479: contig of 6483 bp in length
* 125480 125579: gap of 100 bp
* 125580 130800: contig of 5221 bp in length
* 130801 130900: gap of 100 bp
* 130901 137670: contig of 6770 bp in length
* 137671 137770: gap of 100 bp
* 137771 145779: contig of 8009 bp in length
* 145780 145879: gap of 100 bp
* 145880 152623: contig of 6744 bp in length.
```

```
Query Match 52.4%;  Score 262.2;  DB 2;  Length 152623;
Best Local Similarity 79.4%;  Pred. No. 1.8e-69;
Matches 366;  Conservative 0;  Mismatches 3;  Indels 92;  Gaps 1;

Qy      76 TTACTGAAAGTCAAAAGTGGTGCCGCCCCTPACATCTCTCAAITGTGGTTCGAATAATTACATCA 135
           ||   |||||||| |||||||||||  ||||||| ||||||||||| |||||||||||||||||
Db  120891 TTTAGGAAGTCAAAGTGGTGCGCCCCTACATCTCTCAAATGTGGTTCGAATAATTACATCA 120832

Qy     136 GAGCTCTATCGATCACTGGGAGATGTGCCTCCGTGATGTTGATGCCAAGGCTTTGGTCGC 195
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  120831 GAGCTCTATCGATCACTGGGAGATGTTGCCTCCGTGATGTTGATGCCAAGGCTTTGGTCGC 120772

Qy     196 TCTGACTTTCTTCTGGTGTATGGGGATGCATCTCAAACATCAATAATATCACCAGAGCCCTT 255
           |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  120771 TCTGACTTTCTTCTGGTGTATGGGGATGTCATCTCAAACATCAATAATATCACCAGAGCCCTT 120712

Qy     256 GAGGAACAC----------------------- 264
           ||||||||
Db  120711 GAGGAACACAGGTCAGGATGGGAAAATGACAGGAACAAGGGTTAAAGACCAGCAGAGCCC 120652

Qy     265 ------------------------------AGGTTGAGACGGAAGCTAG 283
                                         |||||||||||||||
Db  120651 TGAGACTGCTTTTTTGCAGTTGTCTGTCCCTCGTCCTTTATAGGTTGAGACGGAAGCTAG 120592

Qy     284 AAAAAAAATGTTTCTGTGATGACGAATGATCTTCAAGGAGTCATCCCCAGCCACCAACTC 343
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  120591 AAAAAAAATGTTTCTGTGATGACGAATGATCTTCAAGGAGTCATCCCCAGCCACCAACTC 120532

Qy     344 GTTGCCACGAAGACAAATGTGGTGTAGTGGCTGTGGAGTAGTACCACAAAACAGGGTTCTCCATT 403
           |||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db  120531 GTTGCCACGAAGACAAATGTGGTTAGTGGCTGTGGAGTAGTACCACAAAACAGGGTTCTCCATT 120472

Qy     404 TTCAGAAGACCCAGGGTCTCCGGCGTTTTGCAITTCCTCTG 444
           |||||||||||||||||||||||||||||||||||||||
Db  120471 TTCAGAAGACCCAGGGTCTCCGGCGTTTTGCAITTCCTCTG 120431
```

RESULT 8
AC048331/c
LOCUS   AC048331   229586 bp   DNA   linear   HTG 31-JAN-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-48102, WORKING DRAFT SEQUENCE,
            23 unordered pieces.
ACCESSION   AC048331
VERSION   AC048331.31 GI:18449516
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 229586)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

COMMENT

TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```
  Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
  Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
  Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
  Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
  Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
  Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
  Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
  Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
  Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
  Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
  Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
  Moser,M., Neal,D., Newton,E., Newson,N., Nguyen,A., Nguyen,N.,
  Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
  Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
  Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
  Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
  Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
  Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
  Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
  Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
  Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
  Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
  Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
  Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
  Worley,K.C.
  Direct Submission
  Submitted (14-APR-2000) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Jan 31, 2002 this sequence version replaced gi:17974902.
```

REFERENCE   2 (bases 1 to 229586)
AUTHORS   Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished

```
-----------------------Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
-----------------------Project Information
            Center project name: HANQ
            Center clone name: RP11-48I02
-----------------------Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-primer Bodipy: 30% of reads
            Chemistry: Dye-terminator Big Dye: 70% of reads
            Assembly program: Phrap; version 0.990329First call to
  findPhrapList
            Consensus quality: 230423 bases at least Q40
            Consensus quality: 245348 bases at least Q30
            Consensus quality: 254396 bases at least Q20
            Estimated insert size: 236955; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; agarose-fp estimation
 -----------------------
 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 102971  102971: contig of 102971 bp in length
 * 102972  103071: gap of unknown length
 * 103072  138452: contig of 35381 bp in length
 * 138453  138552: gap of unknown length
 * 138553  157218: contig of 18666 bp in length
 * 157219  157318: gap of unknown length
 * 157319  165594: contig of 8276 bp in length
 * 165595  165694: gap of unknown length
```

FEATURES   Location/Qualifiers
     source   1..229586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-48I02"

```
 * 165695  170673: contig of 4979 bp in length
 * 170674  170773: gap of unknown length
 * 170774  177654: contig of 6881 bp in length
 * 177655  177754: gap of unknown length
 * 177755  181770: contig of 4016 bp in length
 * 181771  181870: gap of unknown length
 * 181871  185296: contig of 3426 bp in length
 * 185297  185396: gap of unknown length
 * 185397  189950: contig of 4554 bp in length
 * 189951  190050: gap of unknown length
 * 190051  193028: contig of 2978 bp in length
 * 193029  193128: gap of unknown length
 * 193129  195613: contig of 2485 bp in length
 * 195614  195713: gap of unknown length
 * 195714  199136: contig of 3423 bp in length
 * 199137  199236: gap of unknown length
 * 199237  202491: contig of 3255 bp in length
 * 202492  202591: gap of unknown length
 * 202592  205157: contig of 2566 bp in length
 * 205158  205257: gap of unknown length
 * 205258  207869: contig of 2612 bp in length
 * 207870  207969: gap of unknown length
 * 207970  210484: contig of 2515 bp in length
 * 210485  210584: gap of unknown length
 * 210585  213846: contig of 3262 bp in length
 * 213847  213946: gap of unknown length
 * 213947  217298: contig of 3352 bp in length
 * 217299  217398: gap of unknown length
 * 217399  220482: contig of 3084 bp in length
 * 220483  220582: gap of unknown length
 * 220583  223048: contig of 2466 bp in length
 * 223049  223148: gap of unknown length
 * 223149  225284: contig of 2136 bp in length
 * 225285  225384: gap of unknown length
 * 225385  227396: contig of 2012 bp in length
 * 227397  227496: gap of unknown length
 * 227497  229586: contig of 2090 bp in length.
```

BASE COUNT   56229 a   57475 c   54442 g   59219 t   2221 others
ORIGIN

```
Query Match            52.4%;  Score 262.2;  DB 2;  Length 229586;
Best Local Similarity  79.4%;  Pred. No. 1.9e-69;
Matches 366;  Conservative 0;  Mismatches 3;  Indels 92;  Gaps 1;

Qy      76 TTACTGAAGTCAAAGTGGGTGCGCGCCCCTTACACATCTCAATGTGGTTCGAATAATTACATCA 135
           ||||||||||| |||| |||| ||||| ||||||| ||||||||||||||||||||||||||
Db  152415 TTTAGGAAGTCAAAGTGGGTCGCGCCCCTACACATCTCAATGTGGTTCGAATAATTACATCA 152356

Qy     136 GAGGTCTATCGATCACTGGGAGATGTCCTCCGTGATGTTGATCCCAAGGCTTTGGTGCGC 195
           ||||||||||||||| |||||||||||||||||||||||||||||||||| ||||||||
Db  152355 GAGCTCTATCGATCACTGGGAGATGTCCTCCGTGATGTTGATCCCAAGGCTTTGGTGCGC 152296

Qy     196 TCTGACTTTCTTCTGGTGTATGGGGATGTCATCTCAAACATCAATATCACCAGAGCCCTT 255
           ||||||||||||| |||||||| || |||||||||||||||||||||||||||||||||||
Db  152295 TCTGACTTTCTTCTGGTGTATGGGGATGTCATCTCAAACATCAATATCACCAGAGCCCTTT 152236

Qy     256 GAGGAACAC------------------------------------------------ 264
           |||||||||
Db  152235 GAGGAACACAGGTCAGGATGGGAAAATGACAGGAACAAGGGTTAAAGACCAGCAGAGCC 152176

Qy     265 ---------------------------------------------AGGTTGAGACGCGAAAGCTAG 283
                                                        ||||||||||| ||||||||
Db  152175 TGAGACTGCTTGCTTTTTTGCAGTTCTGTCCCTCCGTTCCTTTATAGGTTGAGACGGAAGCTAG 152116

Qy     284 AAAAAAAATGTTTCTGTGATGACGATGATCTTCAAGGAGTCATCCCCAGCACCCAACTC 343
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  152115 AAAAAAAATGTTTCTGTGATGACGATGATCTTCAAGGAGTCATCCCCAGCACCCAACTC 152056
```

```
Qy      344 GTTGCCACGAAGACAAATGTGGTAGTGCCTGTGGATAGTACCACAAAACAGGGTTCTCCATT 403
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   152055 GTTGCCACGAAGACAAATGTGGTAGTGCCTGTGGATAGTACCACAAAACAGGGTTCTCCATT 151996

Qy      404 TTCAGAAGACCCAGGGTCTCCGGCGTTTTGCANTTCCTCTG 444
            |||||||||||||||||||||||||||||||||||||||||
Db   151995 TTCAGAAGACCCAGGGTCTCCGGCGTTTTGCANTTCCTCTG 151955
```

RESULT 9
AC092931/9

```
LOCUS       AC092931             298408 bp    DNA     linear   HTG 09-AUG-2001
DEFINITION  Homo sapiens chromosome 3q clone RP11-131020, WORKING DRAFT
            SEQUENCE, 42 unordered pieces.
ACCESSION   AC092931
VERSION     AC092931.1  GI:15135474
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 298408)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            AlsBrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,M., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopolous,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Liew,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,P., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,J., Primus,E., Pu,L.L.,
            Puiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y.F., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 298408)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     Center: Baylor College of Medicine
            Center code: BCM
            Center: Genome Center
```

```
* Web site: http://www.hgsc.bcm.tmc.edu/
* Contact: hgsc-help@bcm.tmc.edu
* ------------- Project Information -------------
* Center project name: HCXV
* Center clone name: RP11-131020
* ------------- Summary Statistics -------------
* Assembly program: Phrap; version 0.990329
* Consensus quality: 276956 bases at least Q40
* Consensus quality: 302270 bases at least Q30
* Consensus quality: 320294 bases at least Q20
* Consensus quality: 320294 bases at least Q20
* Estimated insert size: 310663; sum-of-contigs estimation
* Estimated insert size: 310663; sum-of-contigs estimation
* Quality coverage: 0x in Q20 bases; agarose-fp estimation
* Quality coverage: 3.7x in Q20 bases; sum-of-contigs estimation
*
* NOTE: Estimated insert size may differ from sequence length
*       (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
*       consists of 42 contigs. The true order of the pieces
*       is not known and their order in this sequence record is
*       arbitrary. Gaps between the contigs are represented as
*       runs of N, but the exact sizes of the gaps are unknown.
*       This record will be updated with the finished sequence
*       as soon as it is available and the accession number will
*       be preserved.
*
*          1      33307: contig of 33307 bp in length
*      33308      33407: gap of unknown length
*      33408      55576: contig of 22169 bp in length
*      55577      55676: gap of unknown length
*      55677      71524: contig of 15848 bp in length
*      71525      71624: gap of unknown length
*      71625      84811: contig of 13187 bp in length
*      84812      84911: gap of unknown length
*      84912      96841: contig of 11930 bp in length
*      96842      96941: gap of unknown length
*      96942     107204: contig of 10263 bp in length
*     107205     107304: gap of unknown length
*     107305     119046: contig of 11742 bp in length
*     119047     119146: gap of unknown length
*     119147     128812: contig of 9666 bp in length
*     128813     128912: gap of unknown length
*     128913     139336: contig of 10424 bp in length
*     139337     139436: gap of unknown length
*     139437     148408: contig of 8972 bp in length
*     148409     148508: gap of unknown length
*     148509     156373: contig of 7865 bp in length
*     156374     156473: gap of unknown length
*     156474     165327: contig of 8854 bp in length
*     165328     165427: gap of unknown length
*     165428     173513: contig of 8086 bp in length
*     173514     173613: gap of unknown length
*     173614     180753: contig of 7140 bp in length
*     180754     180853: gap of unknown length
*     180854     187919: contig of 7066 bp in length
*     187920     188019: gap of unknown length
*     188020     195551: contig of 7532 bp in length
*     195552     195651: gap of unknown length
*     195652     202233: contig of 6583 bp in length
*     202234     202334: gap of unknown length
*     202335     208703: contig of 6369 bp in length
*     208704     208803: gap of unknown length
*     208804     216637: contig of 7835 bp in length
*     216638     216738: gap of unknown length
*     216739     221125: contig of 4387 bp in length
*     221126     221225: gap of unknown length
*     221226     225147: contig of 3922 bp in length
*     225148     225247: gap of unknown length
*     225248     230807: contig of 5561 bp in length
*     230808     230907: gap of unknown length
*     230908     236269: contig of 5362 bp in length
*     236270     236370: gap of unknown length
*     236371     240940: contig of 4570 bp in length
*     240941            : gap of unknown length
```

FEATURES
```
source           Location/Qualifiers
                 1..298408
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /chromosome="3q"
                 /clone="RP11-131020"
```

BASE COUNT   77516 a   69113 c   67054 g   80533 t   4192 others
ORIGIN

Best Local Similarity: 77.8%;   Pred. No. 4.1e-66;
Query Match   50.3%;   Score 251.4;   DB 2;   Length 298408;
Matches 360;   Conservative 0;   Mismatches 11;   Indels 92;   Gaps 1;

```
Qy    76  TTACTGAAGTCAAAGTGGGTGCCCGCCCCTACATCTCTCAAGTCGAATAATTACATCA   135
              ||                |||  ||||||||||||||||||||||||||||||
Db 167430  TTTAGGAAGTCAAAAGTGTGCCCCCTACATCTCTCAAGTGGTTCGAATAATTACATCA  167371

Qy   136  GAGCTCTATCGATCACTGGGAGATGTCCTCCGTGATGTTGATGCCAAGGCTTTGGTGCGC   195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 167370  GAGCTCTATCGATCACTGGGAGATGTCCCCGTGATGTTGATGCCAAGGCTTTGGTGCGC   167311

Qy   196  TCTGACTTTCTTCTGGGTGTATGGGGATGTCATCTCAAACATCAATATCACCAGAGCCCTT   255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 167310  TCTGACTTTCTTCTGGGTGTATGGGGATGTCATCTCAAACATCAATATCACCAGAGCCCTT   167251

Qy   256  GAGGAACAC-----------   264
              ||||||||
Db 167250  GAGGAACAGAGTCAGATGGGAAAAATGACAGGAACAACAAGGGTTAAAGACCAGCAGAGCCC   167191

Qy   265  -----------   265
              |||||||||||
Db 167190  TGAGACTGCTTTTTTGCAGATCTGTCCCTCGCTTTTATAGGTTGAGACGGAAGCTAG   167131

Qy   284  AAAAAAATGTTTCTGTGATGACGGATGATCTTCAAGGAGTCATCCCCACCCCAAACTC   343
              |||||||||||||||||||||||||||||||||||||||||||||
Db 167130  AAAAAAATGTGTCTGTGATGACGATGATCTTCAAGGAGTCATCCCCAGCACCCCAGCCACCCCAACTC   167071

Qy   344  GTTGCCACGAGACAATGTGGTAGTGGCTGTGTTGGATAGTAGTACCACAAACACAGGGTTCTCCCATT   403
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 167070  GTTGCCACAAAAACACTGTGTAGTGGCTGTGAGTGCGTGGGATAGTACCACACAACCAGGGTTCTCCATT   167011
```

RESULT 10

Qy   404 TTCAGAAAAACCACGGGGCTCCGGCGTTTTGCATTCCTCTGGG   446
              ||||||||||||||||||||||||||||||||||||||||
Db 167010 TTCAGAAAACCACGGGGCTCCGGCGTTTTGCATTCCTCTGGG   166968

LOCUS   AC074027   199359 bp   DNA   linear   HTG 25-JAN-2002
DEFINITION Mus musculus chromosome 16 clone RP23-113H11 strain C57BL/6/J,
            WORKING DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION   AC074027
VERSION   AC074027.7 GI:12313746
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 199359)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE   High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 199359)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-JUL-2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
COMMENT   On Jan 19, 2001 this sequence version replaced gi:12043571.
            -------------------- Genome Center
            Center: Harvard Partners Genome Center
            Center Code: HPGC
            Web site:
            http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous
            e.html
            Contact: gnktm@capecod.bwh.harvard.edu
            -------------------- Summary Statistics
            Center project name: ADH
            Sequencing vector: pUC18; L08752
            Chemistry: Dye-terminator Big Dye; 100%
            *Consensus quality: 184639 at least Q20
            *Consensus quality: 178829 at least Q30
            *Consensus quality: 168661 at least Q40
            Estimated insert size: 198699 - sum-of-contigs
            **Estimated insert size: agarose-FP - N/A
            Quality coverage: agarose-FP - N/A
            Quality coverage: 4.8 x in Q20 bases; sum-of-contigs estimation
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1: contig of 12682 bp in length
            *   12683: gap of unknown length
            *   12702: contig of 12682 bp in length
            *   12703: gap of unknown length
            *   23292: contig of 10590 bp in length
            *   23293: gap of unknown length
            *   23312: contig of 15780 bp in length
            *   23313: gap of unknown length
            *   39092: contig of unknown length
            *   39093: gap of unknown length
            *   39112: contig of 11996 bp in length
            *   39113: gap of unknown length
            *   51108: contig of unknown length
            *   51109: gap of unknown length
            *   51128: contig of unknown length
            *   51129: gap of unknown length
            *   60218: contig of 9090 bp in length
            *   60219: gap of unknown length
            *   60238: contig of unknown length
            *   60239: gap of unknown length
            *   70060: contig of 9822 bp in length
```
            *         241041: gap of unknown length
            *         244908: contig of 3868 bp in length
            *         244909: gap of unknown length
            *         245008: contig of 4133 bp in length
            *         245009: gap of unknown length
            *         249141: contig of 2767 bp in length
            *         249142: gap of unknown length
            *         249242: contig of 4949 bp in length
            *         249291: gap of unknown length
            *         254190: contig of 3451 bp in length
            *         254291: gap of unknown length
            *         254742: contig of 2840 bp in length
            *         257741: gap of unknown length
            *         257841: contig of unknown length
            *         257842: gap of unknown length
            *         260681: contig of 2840 bp in length
            *         260682: gap of unknown length
            *         260782: contig of 2767 bp in length
            *         263548: gap of unknown length
            *         263549: contig of 2989 bp in length
            *         263648: gap of unknown length
            *         263649: contig of 3021 bp in length
            *         266637: gap of unknown length
            *         266638: contig of 3021 bp in length
            *         266737: gap of unknown length
            *         266738: contig of 3272 bp in length
            *         269758: gap of unknown length
            *         269759: contig of 3272 bp in length
            *         269858: gap of unknown length
            *         269859: contig of unknown length
            *         273130: gap of unknown length
            *         273131: contig of 2739 bp in length
            *         273230: gap of unknown length
            *         273231: contig of 2739 bp in length
            *         275969: gap of unknown length
            *         275970: contig of 4251 bp in length
            *         276069: gap of unknown length
            *         276070: contig of 4251 bp in length
            *         280320: gap of unknown length
            *         280321: contig of 2264 bp in length
            *         280420: gap of unknown length
            *         280421: contig of 2264 bp in length
            *         282684: gap of unknown length
            *         282685: contig of 2926 bp in length
            *         282784: gap of unknown length
            *         282785: contig of 2926 bp in length
            *         285710: gap of unknown length
            *         285711: contig of 2926 bp in length
            *         285810: gap of unknown length
            *         285811: contig of 2926 bp in length
            *         288736: gap of unknown length
            *         288737: contig of 2049 bp in length
            *         288836: gap of unknown length
            *         288837: contig of 2049 bp in length
            *         290885: gap of unknown length
            *         290886: contig of 3175 bp in length
            *         290985: gap of unknown length
            *         290986: contig of 3175 bp in length
            *         294160: gap of unknown length
            *         294161: contig of 2017 bp in length
            *         294260: gap of unknown length
            *         294261: contig of 2017 bp in length
            *         296277: gap of unknown length
            *         296278: contig of 2031 bp in length
            *         296377: gap of unknown length
            *         296378: contig of 2031 bp in length
            *         298408: contig of 2031 bp in length.
```

```
FEATURES             Location/Qualifiers
source               1..199359
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-113H11"
                     /sex="male"
misc_feature         1..12682
                     /note="assembly_name:Contig91"
misc_feature         12703..23292
                     /note="assembly_name:Contig90"
misc_feature         23313..39092
                     /note="assembly_name:Contig89"
misc_feature         39113..51108
                     /note="assembly_name:Contig89"
misc_feature         51129..60218
                     /note="assembly_name:Contig88"
misc_feature         60239..70060
                     /note="assembly_name:Contig87"
misc_feature         70081..79260
                     /note="assembly_name:Contig86"
misc_feature         79281..90181
                     /note="assembly_name:Contig85"
misc_feature         90202..98036
                     /note="assembly_name:Contig84"
misc_feature         98057..105849
                     /note="assembly_name:Contig83"
misc_feature         105870..114552
                     /note="assembly_name:Contig82"
misc_feature         114573..123701
                     /note="assembly_name:Contig81"
misc_feature         123722..131069
                     /note="assembly_name:Contig80"
misc_feature         131090..138295
                     /note="assembly_name:Contig79"
misc_feature         138316..144002
                     /note="assembly_name:Contig78"
misc_feature         144023..149683
                     /note="assembly_name:Contig77"
misc_feature         149704..154301
                     /note="assembly_name:Contig76"
misc_feature         154322..158504
                     /note="assembly_name:Contig75"
misc_feature         158525..161217
                     /note="assembly_name:Contig74"
misc_feature         161238..164558
                     /note="assembly_name:Contig73"
misc_feature         164579..167331
                     /note="assembly_name:Contig72"
                     /vector_side:left
                     /clone_end:T7
misc_feature         167352..171444
                     /note="assembly_name:Contig71"
                     /vector_side:right
                     /clone_end:SP6
misc_feature         171465..175319
                     /note="assembly_name:Contig70"
misc_feature         175340..178350
                     /note="assembly_name:Contig69"
misc_feature         178371..181271
                     /note="assembly_name:Contig68"
misc_feature         181292..183851
                     /note="assembly_name:Contig67"
misc_feature         183872..187525
                     /note="assembly_name:Contig66"
misc_feature         187546..189662
                     /note="assembly_name:Contig65"
misc_feature         189683..190563
                     /note="assembly_name:Contig64"
misc_feature         190584..192520
                     /note="assembly_name:Contig63"
misc_feature         192541..194141
                     /note="assembly_name:Contig62"
misc_feature         194162..196034
                     /note="assembly_name:Contig61"
misc_feature         196055..198800
                     /note="assembly_name:Contig60"
misc_feature         198821..199359
                     /note="assembly_name:Contig59"
misc_feature         199359
                     /note="assembly_name:Contig58"

*    70061   70080: gap of unknown length
*    70081   79260: contig of 9180 bp in length
*    79261   79280: gap of unknown length
*    79281   90181: contig of 10901 bp in length
*    90182   90201: gap of unknown length
*    90202   98036: contig of 7835 bp in length
*    98037   98056: gap of unknown length
*    98057  105849: contig of 7793 bp in length
*   105850  105869: gap of unknown length
*   105870  114552: contig of 8683 bp in length
*   114553  114572: gap of unknown length
*   114573  123701: contig of 9129 bp in length
*   123702  123721: gap of unknown length
*   123722  131069: contig of 7348 bp in length
*   131070  131089: gap of unknown length
*   131090  138295: contig of 7206 bp in length
*   138296  138315: gap of unknown length
*   138316  144002: contig of 5687 bp in length
*   144003  144022: gap of unknown length
*   144023  149683: contig of 5661 bp in length
*   149684  149703: gap of unknown length
*   149704  154301: contig of 4598 bp in length
*   154302  154321: gap of unknown length
*   154322  158504: contig of 4183 bp in length
*   158505  158524: gap of unknown length
*   158525  161217: contig of 2693 bp in length
*   161218  161237: gap of unknown length
*   161238  164558: contig of 3321 bp in length
*   164559  164578: gap of unknown length
*   164579  167331: contig of 2753 bp in length
*   167332  167351: gap of unknown length
*   167352  171444: contig of 4093 bp in length
*   171445  171464: gap of unknown length
*   171465  175319: contig of 3855 bp in length
*   175320  175339: gap of unknown length
*   175340  178350: contig of 3011 bp in length
*   178351  178370: gap of unknown length
*   178371  181271: contig of 2901 bp in length
*   181272  181291: gap of unknown length
*   181292  183851: contig of 2560 bp in length
*   183852  183871: gap of unknown length
*   183872  187525: contig of 3654 bp in length
*   187526  187545: gap of unknown length
*   187546  189662: contig of 2117 bp in length
*   189663  189682: gap of unknown length
*   189683  190563: contig of 881 bp in length
*   190564  190583: gap of unknown length
*   190584  192520: contig of 1937 bp in length
*   192521  192540: gap of unknown length
*   192541  194141: contig of 1601 bp in length
*   194142  194161: gap of unknown length
*   194162  196034: contig of 1873 bp in length
*   196035  196054: gap of unknown length
*   196055  198800: contig of 2746 bp in length
*   198801  198820: gap of unknown length
*   198821  199359: contig of 539 bp in length.
```

```
BASE COUNT     52621 a   47137 c   46866 g   51719 t   1016 others

Query Match
Best Local Similarity  36.6%;       Score 182.8;   DB 2;   Length 199359;
                       69.1%;       Pred. No. 5.7e-45;
Matches 307;  Conservative 0;  Mismatches 57;  Indels 80;  Gaps 1;

Qy     81 GAGTCAAAGTGGTGCCGCCCTACATCTCTCAAGTGGTTCGAATAATTACATCAGAGCT 140
          ||| || ||||||||||||  || ||||||| | |||| |||||| ||||||||||||
Db 139094 GAAATCCAAGTGGTGCCATCCTACGTCCCCACGTGTTGTTCGGATAATCACATCAGAGCT 139153
```

```
RESULT 11

LOCUS       AC073791/c      239574 bp    DNA    linear    HTG 18-JUL-2000
DEFINITION  Mus musculus clone RP23-419H17, WORKING DRAFT SEQUENCE, 15 ordered
            pieces.
ACCESSION   AC073791
VERSION     AC073791.2  GI:9256800
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 239574)
AUTHORS     DoE Joint Genome Institute.
TITLE       Sequencing of Mouse.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 239574)
AUTHORS     DoE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (29-JUN-2000) Production Sequencing Facility, DoE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:8810408.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----Project Information
            Center Project Name: 1897233
            Center Project Name:
            Center clone name: RPCI-23_419H17
            -----
            Summary Statistics
            Consensus quality: 230784 bases at least Q40
            Consensus quality: 236302 bases at least Q30
            Consensus quality: 237537 bases at least Q20
            Estimated insert size: 218000; agarose-fp estimation
            Estimated insert size: 238924; sum-of-contigs estimation
            Quality coverage: 10.77 in Q20 bases; agarose-fp estimation
            Quality coverage: 9.83 in Q20 bases; sum-of-contigs estimation
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.

FEATURES             Location/Qualifiers
     source          1..239574
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="RPCI mouse BAC library 23"
                     /clone="RP23-419H17"

          1   49594: contig of 49594 bp in length
      49595   49694: gap of unknown length
      49695   60237: contig of 10543 bp in length
      60238   60337: gap of unknown length
      60338   62291: contig of 1954 bp in length
      62292   62391: gap of unknown length
      62392   95540: contig of 33149 bp in length
      95541   95640: gap of unknown length
      95641  100286: contig of 4646 bp in length
     100287  100386: gap of unknown length
     100387  111056: contig of 10670 bp in length
     111057  111156: gap of unknown length
     111157  118803: contig of 7647 bp in length
     118804  118903: gap of unknown length
     118904  138921: contig of 20018 bp in length
     138922  139021: gap of unknown length
     139022  165699: contig of 26678 bp in length
     165700  165799: gap of unknown length
     165800  167477: contig of 1678 bp in length
     167478  167577: gap of unknown length
     167578  179458: contig of 11881 bp in length
     179459  179558: gap of unknown length
     179559  190231: contig of 10673 bp in length
     190232  190331: gap of unknown length
     190332  222927: contig of 32596 bp in length
     222928  223027: gap of unknown length
     223028  234672: contig of 11645 bp in length
     234673  234772: gap of unknown length
     234773  239574: contig of 4802 bp in length

BASE COUNT    64477 a  52869 c  55524 g  65301 t  1403 others
ORIGIN      1..239574
```

```
Query Match       36.6%;   Score 182.8;  DB 2;  Length 239574;
Best Local Similarity  69.1%;   Pred. No. 5.8e-45;
Matches 307;   Conservative  0;  Mismatches  57;  Indels  80;  Gaps  1;

Qy   81  GAAGTCAAAGTGGTGCCGCCCTACACTCTCTCAATGTGGTTCGAATAATTAACATCAGAGCT  140
         ||||| ||||||||||| ||||||| || | ||| ||||||||| |||| |  |||||||||
Db 93430 GAAATCCAAGTGGTGCCATCCTACGTCCCCCAACGGTTGTTCGGATAATCACATCAGAGCT 93371

Qy  141  CTATCGATCACTGGGAGATGTCTCCGTGATGTTGATGCCAAGGCTTTGGTGTGCGCTCTGA  200
         |||||||||||| |||||| |||| ||||| |||||||||||||| ||||||| ||||||||
Db 93370 ATACCGATCACTAGGAGAGACGTCCTCCGACGTCGATGCCAAGGCCTTGGTGCGCCTCTGA 93311

Qy  201  CTTTCTTCTTGGTGTATGGGGATGTCATCTCAAACATCAATATCACCAGCCCCTTGAGGA  260
         ||||| |||||| |||| ||| ||||||||||||| |||||||||  | || |||| |||||
Db 93310 CTTTCTCTTGATATATGGAGGATGTCATCTCAAATATCAATATCTGCAGACCCCTGGAGGA 93251

Qy  261  ACAC------                                              264
         ||||
Db 93250 ACACAGGTCAGGCTGGGGAAGACTGCAGGACAGACTTAGGGACCAGCAGAGTCCTAAGA 93191

Qy  265  ----------AGGTTGAGGACGGAAGCTAGAGAAAAAAATGTTTTCTGTG  300
                   ||||| || |||||||||||||||| |||||||||| ||||
Db 93190 CTAACGTGCC--AGGTTGAGACGGAAGCTAGAAGAAAGTTGAGAAAAAAATGTTTCTGTG 93131

Qy  301  ATGACGATGAATCTTCAAGGAGTCATCCCCAGCCACCCAACTCGTTGCCACGAAGACAAT  360
         ||||||| |||| ||||||| |||| || |||||||||  |||||||||| |||||||||
Db 93130 ATGACAAATGGTCTTCAAAGAGTCGTCACCCAGCCACCCTACCCAGCCAACTCGTTGCCACGAAGACAAC 93071

Qy  361  GTGGTAGTTGGGCTGTGGATTAGTACCACAAACAGGGTGTTCTCCATTTTCAGAAGACCCAGGGT  420
         |||||| ||||||||||| |  |  ||||| ||||||||||||| |  ||||||||||||| |||
Db 93070 GTGGTGATTGGGCTGTGGGACAGCGCCACCACAACAGGGGTTCTTCACTTCCAGAAGACCCAAGGC 93011
```

```
Qy    421 CTCCGGCGGCGTTTTGCATTTCCTCTG 444
          |||||||| ||| |  | |||||| ||
Db  93010 CTCCGGCGGCTTTTTCCTTCCATTG 92987
```

```
RESULT 12
LOCUS       RNU19511               13164 bp    DNA     linear   ROD 30-OCT-1996
DEFINITION  Rattus norvegicus initiation factor-2Be gene, complete cds.
ACCESSION   U19511
VERSION     U19511.1  GI:924596
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 13164)
  AUTHORS   Flowers,K.M., Mellor,H., Matts,R.L., Kimball,S.R. and
            Jefferson,L.S.
  TITLE     Cloning and characterization of complementary and genomic DNAs
            encoding the epsilon-subunit of rat translation initiation
            factor-2B
  JOURNAL   Biochim. Biophys. Acta 1307 (3), 318-324 (1996)
  MEDLINE   96305355
REFERENCE   2  (bases 1 to 13164)
  AUTHORS   Mellor,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1995) Harry Mellor, Cellular Molecular
            Physiology, Penn State College of Medicine, Milton S. Hershey
            Medical Center, Hershey, PA 17033, USA
FEATURES             Location/Qualifiers
     source          1..13164
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /clone_lib="rat p1 genomic"
     promoter        383..3496
                     /note="TATA-like/a-Pal element"
     mRNA            join(3512..3790,4811..4935,5779..5964,6044..6221,
                     6310..6390,6646..6723,6961..7273,8857..9002,9241..9382,
                     9504..9605,10187..10294,10439..10529,11066..11189,
                     11633..11758,12343..12453,12571..12932)
                     /product="initiation factor-2Be"
     exon            3512..3790
                     /number=1
     CDS             join(3611..3790,4811..4935,5779..5964,6044..6221,
                     6310..6390,6646..6723,6961..7273,8857..9002,9241..9382,
                     9504..9605,10187..10294,10439..10529,11066..11189,
                     11633..11758,12343..12453,12571..12630)
                     /codon_start=1
                     /product="initiation factor-2Be"
                     /protein_id="AAB17691.1"
                     /db_xref="GI:924597"
                     /translation="MAATAAVPSAVGGRANKRGGSGSGGGGTGGAEEEPPPLQAVLVA
                     DSFDRFFISKDQPRVLLPLANVALIDYTLEFLATGVQETFVFCCWKAAQIKEHLQ
                     KSKWCHPTSLNVVRITTISDLYIRSLGDVLRDVDAKALVRSDELLIYGDVVSNISKAL
                     EEHRLRRKLEKNVSVMTMVFKESSPSHPTRCHEDNVVLAVDSTTNRILHFQKTQGLRH
                     FSFPLGLFQGSLDGVEIRIVDLDCHISICSPOVAQLFTDNFDYQTRDDFVRGLLVNEE
                     ILGNQIHLHVTSREYGSRVSNLHMYSAVCTDVIRRWYPLTPEVNFTDSSTQSYTHSR
                     HNIYRGPEVSLGHGSVLEENVLLGAGTVVGSNCSITNSVIGPNCHLGDNVVLDQAYLM
                     QGVRVAAGAQIHQSLLCDRAEVKERVILKPHCVLTSQVVVGPDIILPEGSVISLHPPD
                     AEEDEDDGGPSDSDGADQEREKVKLKGYNAEVGPEGQGYLWKAEDVDEKEDEELRQS
                     LWGLMINMEEEESETESERSVDPEELDSRAGSPQLDDIRVFONEVLGTLQRGREENISC
                     DNLVLEINSLKYAYNISLKEVMOVLSHVVLEFPLCQOVDGVLDPNRYCALLPLLKAWS
                     PVFRNYIKRAADHLEALAAIEDFFLEHETIVPSLAKVLMAFYQLEILAEETILSWFSQ
                     RDITDKGQOLRKNQOLQRFIOWLREAEESSDDD."
     exon            4811..4935
                     /number=2
     exon            5779..5964
                     /number=3
     exon            6044..6221
```

```
                     /number=4
     exon            6310..6390
                     /number=5
     exon            6646..6723
                     /number=6
     exon            6961..7273
                     /number=7
     exon            8857..9002
                     /number=8
     exon            9241..9382
                     /number=9
     exon            9504..9605
                     /number=10
     exon            10187..10294
                     /number=11
     exon            10439..10529
                     /number=12
     exon            11066..11189
                     /number=13
     exon            11633..11758
                     /number=14
     exon            12343..12453
                     /number=15
     exon            12571..12932
                     /number=16
BASE COUNT     3257 a   2941 c   3250 g   3716 t
ORIGIN
```

```
Query Match          36.1%;  Score 180.6;  DB 10;  Length 13164;
Best Local Similarity  68.8%;  Pred. No. 2.1e-44;
Matches  305;  Conservative  0;  Mismatches  59;  Indels  79;  Gaps  1;

Qy     81 GAAGTCAAAGTWGGTGCGCCGCCCTACACATCTCAAATGTGGTTCGAATAATTACATCAGAGACTT 140
          || ||  |||  |  | | ||||| |||||||  |||| || | ||||| |||| |||| |||||||
Db   5779 GAAATCCAAAATGGTGCCATCCTTACGTCCCTCAATGTTGTTCGAATAACTACATCAGACTT 5838

Qy    141 CTATCGATCACTGGGAGGATGTCCTCCGTGATGTTGATGCCAAGGCTTTGGTGCGGCTCTGA 200
          || |||| ||| ||  | | ||||| |||| |||||||| |||  ||| |||||| |||||||
Db   5839 GTACCGATCACTCAGGAGAGATGTCCTCCGTGATGTTGATGCCAAGGCCTTGGTGCGGCTCTGA 5898

Qy    201 CTTTCTTTCTGGTGTATGGGGATGTCAYCTCAAACATCAAATATCACCAGAGCCCTTGAGGA 260
          ||| |||||||| || | |||||||| || ||||| |||||||| |||  ||| || |||||
Db   5899 CTTCCTCCCTGATFATATGGGAGGATGTCGYCTCGAATATCAAATATTTCCAAAGCCCTGGAGGA 5958

Qy    261 ACAC----------------- 264
          ||||
Db   5959 ACACAGGTCAGGTGGGAAGACGGCAGGAGGACAGACTTAGGGACCAGTTGAGTCTTAAGAC 6018

Qy    265 ------------------------AGGTTGAGACGGAAGGCTAGAAAAAAATGTTTCTGTGA 301
                                   |||| || | ||| | || | ||||||||||||| ||||
Db   6019 TAAACTCTGCCCCTCCTTCTTATAGGTTAAGGAAGGCTAGAAAAAAATGTGTTCTGTCA 6078

Qy    302 TGACGATGATCTTCAAGGAGTCATCCCCCAGCCACCCAAACTCGTTGCCACGGAGCAAATG 361
          || || ||| ||| || |||| | ||| | | | ||| ||||| |||| ||||| | ||| ||
Db   6079 TGACAATGGTGTTCAAAGAGTCCTCACCCTACACGCGTGCCATGAGGACAATG 6138

Qy    362 TGGTAGTGGGCGTGGGATTAGTACCACAAACAGGGGTTCTCCATTTTCAGAAGAGACCAGGGTC 421
          ||||| ||| |  |||||  |||| ||||| |||| |||| ||| | |||| || || |||
Db   6139 TGGTGCTAGCTGTGGATAGCACCACAAACAGGAGTTCTTCACTTCCAGAAGACCCAAGGCC 6198

Qy    422 TCCGGCGGTTTGCATTTCCTCTG 444
          |||| || | |  |||| ||||| ||
Db   6199 TCCGCCCATTTCTCCTTTCCTTTG 6221
```

```
RESULT 13
LOCUS       AK056400                2542 bp    mRNA    linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ31838 fis, clone NT2RP7000076, weakly similar
            to TRANSLATION INITIATION FACTOR EIF-2B EPSILON SUBUNIT.
ACCESSION   AK056400
VERSION     AK056400.1  GI:16551796
KEYWORDS    oligo capping; fis (full insert sequence).
```

```
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP7 clone:NT2RP7000076.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2542)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2542
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RP7000076"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP7"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 5-weeks retinoic acid (RA)
                     induction."
BASE COUNT      624 a      608 c      651 g      659 t
ORIGIN

RESULT 14
AC074027/c
LOCUS       AC074027              199359 bp    DNA     linear   HTG 25-JAN-2002
DEFINITION  Mus musculus chromosome 16 clone RP23-113H11 strain C57BL/6J,
            WORKING DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION   AC074027
VERSION     AC074027.7  GI:12313746
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            house mouse.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 199359)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.


Query Match              16.3%;
Best Local Similarity   98.8%;
Matches        82;  Conservative   0;  Mismatches   1;  Indels     0;  Gaps     0;

Qy        1 GAATTCCTGACTGCCCACAGGTGTACAGGAAACATTTGTCTTTTGTTCTTTGGCTGGAAAGCTGCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2224 GAATTCCTGACTGCCCACAGGTGTACAGGAAACATTTGTCTTTTGTTCTTTGGCTGGAAAGCTGCT 2283

Qy       61 CAAATCAAAGAAGACATTTACTGAA 83
            ||||||||||||||||||||||||| |
Db     2284 CAAATCAAAGAAGACATTTACTGTA 2306
```

```
FEATURES             Location/Qualifiers
     source          1..199359
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-113H11"
                     /sex="male"
     misc_feature    1..12682
                     /note="assembly_name:Contig91"
     misc_feature    12703..23292
                     /note="assembly_name:Contig90"
     misc_feature    23313..39092
                     /note="assembly_name:Contig89"
     misc_feature    39113..51108
                     /note="assembly_name:Contig88"
     misc_feature    51129..60218
                     /note="assembly_name:Contig87"
     misc_feature    60239..70060
                     /note="assembly_name:Contig86"
     misc_feature    70081..79260
                     /note="assembly_name:Contig85"
     misc_feature    79281..90181
                     /note="assembly_name:Contig84"
     misc_feature    90202..98036
                     /note="assembly_name:Contig83"
     misc_feature    98057..105849
                     /note="assembly_name:Contig82"
     misc_feature    105870..114552
                     /note="assembly_name:Contig81"
     misc_feature    114573..123701
                     /note="assembly_name:Contig80"
     misc_feature    123722..131069
                     /note="assembly_name:Contig79"
     misc_feature    131090..138295
                     /note="assembly_name:Contig78"
     misc_feature    138316..144002
                     /note="assembly_name:Contig77"
     misc_feature    144023..149683
                     /note="assembly_name:Contig76"
     misc_feature    149704..154301
                     /note="assembly_name:Contig75"
```

Gap / contig listing:

```
* 158505   158524:  gap of unknown length
* 158525   161217:  contig of 2693 bp in length
* 161218   161237:  gap of unknown length
* 161238   164558:  contig of 3321 bp in length
* 164559   164578:  gap of unknown length
* 164579   167331:  contig of 2753 bp in length
* 167332   167351:  gap of unknown length
* 167352   171444:  contig of 4093 bp in length
* 171445   171464:  gap of unknown length
* 171465   175319:  contig of 3855 bp in length
* 175320   175339:  gap of unknown length
* 175340   178350:  contig of 3011 bp in length
* 178351   178370:  gap of unknown length
* 178371   181271:  contig of 2901 bp in length
* 181272   181291:  gap of unknown length
* 181292   183851:  contig of 2560 bp in length
* 183852   183871:  gap of unknown length
* 183872   187525:  contig of 3654 bp in length
* 187526   187545:  gap of unknown length
* 187546   189662:  contig of 2117 bp in length
* 189663   189682:  gap of unknown length
* 189683   190563:  contig of 881 bp in length
* 190564   190583:  gap of unknown length
* 190584   192520:  contig of 1937 bp in length
* 192521   192540:  gap of unknown length
* 192541   194141:  contig of 1601 bp in length
* 194142   194161:  gap of unknown length
* 194162   196034:  contig of 1873 bp in length
* 196035   196054:  gap of unknown length
* 196055   198800:  contig of 2746 bp in length
* 198801   198820:  gap of unknown length
* 198821   199359:  contig of 539 bp in length.
```

```
     misc_feature    154322..158504
                     /note="assembly_name:Contig74"
     misc_feature    158525..161217
                     /note="assembly_name:Contig73"
     misc_feature    161238..164558
                     /note="assembly_name:Contig72"
     misc_feature    164579..167331
                     /note="assembly_name:Contig71
                     clone_end:T7
                     vector_side:left"
                     clone_end:SP6
                     vector_side:right"
     misc_feature    167352..171444
                     /note="assembly_name:Contig70"
     misc_feature    171465..175319
                     /note="assembly_name:Contig69"
     misc_feature    175340..178350
                     /note="assembly_name:Contig68"
     misc_feature    178371..181271
                     /note="assembly_name:Contig67"
     misc_feature    181292..183851
                     /note="assembly_name:Contig66"
     misc_feature    183872..187525
                     /note="assembly_name:Contig65"
     misc_feature    187546..189662
                     /note="assembly_name:Contig64"
     misc_feature    189683..190563
                     /note="assembly_name:Contig63"
     misc_feature    190584..192520
                     /note="assembly_name:Contig62"
     misc_feature    192541..194141
                     /note="assembly_name:Contig61"
     misc_feature    194162..196034
                     /note="assembly_name:Contig60"
     misc_feature    196055..198800
                     /note="assembly_name:Contig59"
     misc_feature    198821..199359
                     /note="assembly_name:Contig58"
BASE COUNT    52621 a    47137 c    46866 g    51719 t    1016 others
```

```
RESULT 15

Qy          1  GAATTCCTGACTGCCACAGGTGTACAGGAAAACATTTGTCTTTTTGTTCCTTGGAAAGCTGCT  60
               ||||||||||||||||| |||||| |||||||||||||||  |||||||| ||  || ||||||
Db     144315  GAGTTCCTGACTGCCACGGGTGTACAGGAAAACATTTGTTTCTTTTTGCTCGCTGGGAAAGCTGCT  144256

Qy         61  CAAATCAAAGAACATTTACTCGAA  83
               || ||||||| ||||| |||| |
Db     144255  CAGATCAAAAGAACACTTACAGTA  144233

Query Match         14.4%;  Score 71.8;  DB 2;  Length 199359;
Best Local Similarity   91.6%;  Pred. Nr. 9.5e-11;
Matches   76;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

LOCUS       YSCGCDA                 2903 bp    DNA     linear   PLN 27-APR-1995
DEFINITION  Saccharomyces cerevisiae tRNA, Ile gene, partial cds; guanine
            nucleotide exchange factor subunit gene, complete cds.
ACCESSION   L07115
VERSION     L07115.1  GI:171572
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (strain S288C) DNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2903)
  AUTHORS   Bushman,J.L., Asuru,A.I., Matts,R.L. and Hinnebusch,A.G.
  TITLE     Evidence that GCD6 and GCD7, translational regulators of GCN4, are
            subunits of the guanine nucleotide exchange factor for eIF-2 in
            yeast
  JOURNAL   Unpublished (1992)
FEATURES             Location/Qualifiers
```

```
source          1..2903
                /organism="Saccharomyces cerevisiae"
                /strain="S288C"
                /db_xref="taxon:4932"
                /map="IV; right arm"
                <94..>225
CDS             /codon_start=1
                /evidence=experimental
                /product="tRNA, Ile"
                /protein_id="AAA65497.1"
                /db_xref="GI:171573"
                /note="tRNA is not required for GCD6 gene function"
                /translation="ARVAQWLELRAYSNIRFPKFLCQRPFKQAFKSNATVVGSNPTSS
                "
gene            459..2597
                /gene="GCD6"
CDS             459..2597
                /gene="GCD6"
                /standard_name="eIF-2B"
                /note="putative"
                /codon_start=1
                /product="guanine nucleotide exchange factor, eIF-2B,
                delta subunit"
                /protein_id="AAA65498.1"
                /db_xref="GI:171574"
                /translation="MAGKKGQKKSGLGNHGKNSDMDVEDRLQAVVLTDSYETRFMPLIT
                AVKPRCLLPLANVPLIEYTLEFLAKAGVHEVFLICSSHAQINDYIENSKWNLPWSPF
                KITTIMSPEARCGDVMRDLDNRGIITGDFILVSGDVLTINDFSKMLEFHKKMHLQDK
                DHISTMCLSKASTYPKTRITEPAAFVLDKSTSRCIYQDLPLPSSREKTSIQIDPELL
                DNVDEFVIRNDLLIDCRIDICTSHVPLLFIQENFDVQSLRTDFVKGVTSSDILGKHIYAY
                LTDEYAVRVESWQYDTISQDFLGRWCYPLVLDSNIQDDQTYSYESRHIYKEKDVVLA
                QSCKIGKCTAIGSGTKIGEGTKIENSVIGRNCQIGENIRIKNSFIWDCCIIGNNSIID
                HSLIASNATLGSNVRLNDGCIIGFNVKIDDNMDLRNTKISASPLKNAGSRMYDNESN
                EQFDQDLDDQTLAVSIVGDKGVGYIYESEVSDDEDSSTEACKEINTLSNQLDELYLSD
                DSISSATKTKKRITMSVNSIYTDREEIDSEFEDEDFEKEGIATVERAMENNHDLDTA
                LLELNTLRMSMNVYHEVRIATIFALLRRYVHFIATQTLGPKDAVVKVFNQWGLLFKR
                QAFDEEEYIDLMNIIMEKIVEQSFEPDPLILFSALVSLYDNDIIEEDVIYKWWDNVST
                DPRDEVKLTVKWVEWLQNADEESSSEEE"
BASE COUNT    984 a    462 c    587 g    870 t
ORIGIN

Query Match      14.0%;    Score 70.2;    DB 8;    Length 2903;
Best Local Similarity   52.6%;    Pred. No. 1.9e-10;
Matches  153;  Conservative   0;  Mismatches 138;  Indels   0;  Gaps   0;

Qy        1 GAATTCCTGACTGCCACAGGTGTACAGGAGAAACACATTTGTCTTTGTTGCTGGAAAGCTGCT 60
            |||| ||  | | ||| ||  ||  |  ||||   ||||||  || | |||| ||||||| ||
Db      651 GAATTTTTGGCTAAGGCTCGGCGTACATGAAGTTTCTTAATTGCTCCTTCCATGCCAAC 710

Qy       61 CAAATCAAAGAAACAATTTACTGAAGTCAAAGTGGTGCCGCCCTACATCTCTCAAGTGGTT 120
            |||||| |  | | |||| |  | |||| |||   || | | ||| || || ||||| |||
Db      711 CAAATTAATGACTATATTGAGAATTCTAAGTGGAACTTGCCTGGTCTCCATTTAAATT 770

Qy      121 CGAATAAATTACATCAGAGAGCTTCTATCGATCACTGGGAGATGTCCCTCCGTGATGTTGATCGCC 180
            |  || || ||||  | |||| |||  || |  || |||||||  | |||| ||| ||||| |
Db      771 ACCACCATTATGTCTCCAGAAGCTAGATGTACGGGTCGATGTTATGAGAGATCTAGATAAT 830

Qy      181 AAGGCTTTGGTGCGCCTCTGACTTTCTTCGTGTATGGGGATGGTCATCTCAAACATCAAT 240
            |||||    |  | | |||| | |||||| |||| |||||| ||||  |||||||||||
Db      831 AGAGGTATCATTACTGGGAGATTTTATTTTAGTCAGTGGTGATGTTATGACTAACATCGAT 890

Qy      241 ATCACCAGAGCCCTTGAGGAAACACAGGTTGAGACGGAAGCTAGAAAAAAAT 291
            ||| || |  |  ||  | | | |||| || ||  |  |||||  ||| ||
Db      891 TTCAGCAAAATGCTAGAATTTCACAAAAAAATGCATTTGCAAGATAAAGAT 941
```

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                                       (without alignments)
                                       14616.225 Million cell updates/sec

                                       Copyright (c) 1993 - 2002 Compugen Ltd.
                                       GenCore version 5.1.3

Title: us-09-489-101A-10_COPY_1_500
Perfect score: 500
Sequence: 1 atccctccggtttcctca.........aagcgaaagcacttcattt 500

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs., 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*
  1: gb_ba:*
  2: gb_htg:*
  3: gb_in:*
  4: gb_om:*
  5: gb_ov:*
  6: gb_pat:*
  7: gb_ph:*
  8: gb_pl:*
  9: gb_pr:*
 10: gb_ro:*
 11: gb_sts:*
 12: gb_sy:*
 13: gb_un:*
 14: gb_vi:*
 15: em_ba:*
 16: em_fun:*
 17: em_hum:*
 18: em_in:*
 19: em_mu:*
 20: em_om:*
 21: em_or:*
 22: em_ov:*
 23: em_pat:*
 24: em_ph:*
 25: em_pl:*
 26: em_ro:*
 27: em_sts:*
 28: em_un:*
 29: em_vi:*
 30: em_htg_hum:*
 31: em_htg_inv:*
 32: em_htg_other:*
 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	500	100.0	1580	6	AX201589	AX201589 Sequence
2	500	100.0	1580	6	HUMRPDJKA	L07872 Homo sapien
3	432.8	86.6	1581	6	E03234	E03234 cDNA encodi
4	432.8	86.6	3986	10	MMJKRSB	X17459 Mouse mRNA
5	432.8	86.6	3991	10	S63463	S63463 J kappa rec
6	422	84.4	1500	6	I70264	I70264 Sequence 1
7	362.6	72.5	666	6	AR146567	AR146567 Sequence
8	354.4	70.9	707	6	AR146561	AR146561 Sequence
9	348.8	69.8	1336	6	AX053603	AX053603 Sequence
10	330.6	66.1	1085	10	MMPSEUD2	AX201870 Sequence
11	327.2	65.4	373	9	AX201870	AX201870 Sequence
12	320.8	64.2	1599	9	HUMRBPJKX	L34543 Human recom
13	319.2	63.8	1121	9	HUMRBPJKB	L07873 Human Jk-re
14	317.8	63.6	1600	9	HUMRBPJKY	L34544 Human recom
15	317.6	63.5	1057	10	MMPSEUD1	X59130 M.musculus
16	317.4	63.5	1707	10	MUSIGKPRBP	M81871 Mus musculu
17	312.6	62.5	707	6	AX053609	AX053609 Sequence
18	303.2	60.6	5360	9	D14041	D14041 Homo sapien
19	303.2	60.6	5360	9	AX053604	AX053604 Sequence
20	282.8	56.6	768	9	HUMKBF2	L08904 Homo sapien
21	281.8	56.4	625	6	AX053605	AX053605 Sequence
22	281	56.2	625	6	AX053608	AX053608 Sequence
23 c	278.2	55.6	2535	5	U60093	U60093 Xenopus lae
24	238.8	47.8	2319	5	XLU60094	U60094 Xenopus lae
25	221	44.2	815	6	AX053604	AX053604 Sequence
26 c	213.8	42.8	338	9	HUMRBPJKC	L07874 Human Jk-re
27	201.4	40.3	401	9	HUMRBPJKD	L07875 Human Jk-re
28	174	34.8	2523	3	AF085173	AF085173 Ciona int
29	171.4	34.3	538	6	AX321801	AX321801 Sequence
30	169.8	34.0	3022	3	AB003695	AB003695 Halocynth
31 c	166.4	33.3	111211	9	AC093637	AC093637 Homo sapi
32 c	166.4	33.3	2357	9	AL357494	AL357494 Homo sapi
33 c	166.4	33.3	178889	2	AC097109	AC097109 Homo sapi
34 c	166.4	33.3	190000	2	AC006391	AC006391 Homo sapi
35	160.8	32.2	741	6	AX053606	AX053606 Sequence
36 c	146.4	29.3	219629	2	AC084054	AC084054 Mus muscu
37	137.6	27.5	95445	2	AC016175	AC016175 Homo sapi
38	137.6	27.5	208912	2	AC044869	AC044869 Homo sapi
39	127.4	25.5	2640	3	AY069091	AY069091 Drosophil
40	127.4	25.5	2928	3	M94383	M94383 Drosophila
41	124.2	24.8	1884	3	X58393	X58393 D.melanogas
42	120	24.0	2357	10	MMRBPJK	MMRBPJK M.musculus
43	118.4	23.7	123072	2	AC097397	AC097397 Rattus no
44	115.2	23.0	262	11	G43168	G43168 WIAF-1657-S
45	115.2	23.0	262	11	G43169	G43169 WIAF-1658-S

ALIGNMENTS

RESULT 1
AX201589

LOCUS   AX201589   1580 bp   DNA   linear   PAT 30-AUG-2001
DEFINITION Sequence 10 from Patent WO0153349.
ACCESSION   AX201589
VERSION   AX201589.1 GI:15391438
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1580)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE   Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 10 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES   Location/Qualifiers
     source   1..1580
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

```
RESULT 2
LOCUS       HUMRBPJKA       1580 bp    mRNA    linear   PRI 29-JAN-2001
DEFINITION  Homo sapiens recombination signal binding protein (RBPJK) mRNA,
            partial cds.
ACCESSION   L07872
VERSION     L07872.1  GI:190949
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1580)
  AUTHORS   Amakawa,R., Jing,W., Ozawa,K., Matsunami,N., Hamaguchi,Y.,
            Matsuda,F., Kawaichi,M. and Honjo,T.
  TITLE     Human Jk recombination signal binding protein gene (IGKJRB):
            comparison with its mouse homologue
  JOURNAL   Genomics 17 (2), 306-315 (1993)
   MEDLINE  94010923
   PUBMED   8406481
FEATURES             Location/Qualifiers
     source          1..1580
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3q25"
                     /dev_stage="adult"
     gene            1..>1580
                     /gene="RBPJK"
     CDS             1..137
                     /gene="RBPJK"
                     /codon_start=1
                     /product="recombination signal binding protein"
                     /protein_id="AAA60258.1"
                     /db_xref="GI:553639"
                     /translation="MDHTEGLPAEEEPAHASPSPGKGGERPPPKRLTREAMRNYLKERG
                     DQTVILIHAKVAQKSYGNEKRFFCPPPCVYLMGSGWKKKKEQMERDGCSEQESQPCAF
                     IGIGNSDQEMQQLNLEGKNYCTAKTLYISDSDKRKHFIFSVKMFYGNSDDIGVFLSKR
                     IKVISKPSKKKQSLKNADLCIASGTKVALFNRLRSQTVSTRYLHVEGGNFHASSQWG
                     AFFIHLLDDDESEGEEFTVRDVIHYGIHVQGKLVCVSTGMALPRLIIMKVDKHPALLDA
                     DDPVSQLHKCAFYLKDTERMYLCLSQERIIQFQATPCPKEPNKEMINDGASWTIISTD
                     KAEYTFYEGMGPVLAPVTPVPVVESLQLNGGGDVAMLELTGQNFTPNLRWFGDVEAE
                     TMYRCGESMLCVVPDISAFREGWRWVRQPVQVPVTLVRNDGIIYSTSLTFTYTPEPGP
                     RPHCSVAGAILPANSSQVPPNESNTNSEGSYTNASTNSTSVVSSSTATVVS"
     exon            138..176
                     /gene="RBPJK"
                     /number=1
     exon            177..272
                     /gene="RBPJK"
                     /number=2
     exon            273..438
                     /gene="RBPJK"
                     /number=3
     exon            439..613
                     /gene="RBPJK"
                     /number=4
     exon            614..751
                     /gene="RBPJK"
                     /number=5
     exon            752..864
                     /gene="RBPJK"
                     /number=6
     exon            865..1005
                     /gene="RBPJK"
                     /number=7
     exon            1006..1161
                     /gene="RBPJK"
                     /number=8
     exon            1162..1265
                     /gene="RBPJK"
                     /number=9
     exon            1266..1503
                     /gene="RBPJK"
                     /number=10
     exon            1..137
                     /gene="RBPJK"
                     /number=11
```

```
BASE COUNT    481 a     352 c    359 g    388 t
ORIGIN
Query Match       100.0%;  Score 500; DB 9; Length 1580;
Best Local Similarity  100.0%;  Pred. No. 4.2e-133;
Matches  500;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 ATCCCCTTCCGGTTTTTCCTCAGTCTCCACGTACGTCCCTCAAAAGCGCGTCCTAAAACCCCGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCCCCTTCCGGTTTTTCCTCAGTCTCCACGTACGTCCCTCAAAAGCGCGTCCTAAAACCCCGG 60

Qy     61 ATAACCGGAGCGCGCTCCCCATGGACCACACGGACGAGGGCTTGCCCCGCCGCCGCCGCCTGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATAACCGGAGCGCGCTCCCCATGGACCACACGGACGAGGGCTTGCCCCGCCGCCGCCGCCTGCG 120

Qy    121 CATGCTCCATCGGCCTGGGAAAATTTGGTGAGCGGCGCTCCACCTAAACGACTTACTTAGGGAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CATGCTCCATCGGCCTGGGAAAATTTGGTGAGCGGCGCTCCACCTAAACGACTTACTTAGGGAA 180

Qy    181 GCTATCGCGAAATTATTTAAAAGAGCGAGGGGATCAAAACAGTACTTAATTCTTCATGCAAAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTATCGCGAAATTATTTAAAAGAGCGAGGGGATCAAAACAGTACTTAATTCTTCATGCAAAA 180

Qy    241 GTTGCACAGAAGTCATATGGAAATGAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTGTATAT 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTGCACAGAAGTCATATGGAAATGAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTGTATAT 240
```

RESULT 3

```
LOCUS       E03234      1581 bp    RNA        linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding Rodent Jk rearrangement signal sequence binding
            protein.
ACCESSION   E03234
VERSION     E03234.1  GI:2171451
KEYWORDS    unidentified.
SOURCE      unidentified.
  ORGANISM  unclassified.
REFERENCE   1  (bases 1 to 1581)
  AUTHORS   Honshiyo.Y., Kawaichi.M., Hamaguchi.Y., Matsunami.N., Takahashi.M.,
            Shibayama.S., Nagakuro.K., Yamamoto.Y., Kuze.K., Sagawa.K. and
            Matsuo.T.
  TITLE     JK RECOMBINANT SIGNAL SEQUENCE-BINDING PROTEIN
  JOURNAL   Patent: JP 19912772283-A 1 09-DEC-1991;
            TOYOBO CO LTD
COMMENT     OS  Rodentia
            PN  JP 19912772283-A/1
            PD  09-DEC-1991
            PF  26-MAR-1990 JP 199007317
            PI  HONSHIYO YUU, KAWAICHI MASASHI, HAMAGUCHI YASUJI, PI
            MATSUNAMI NORISADA,
            PI  TAKAHASHI MASAZUMI, SHIBAYAMA SHIRO, NAGAKURO KAZUO, PI
            YAMAMOTO YOSHIKI,
            PI  KUZE KOUGO, SAGAWA KENJI, MATSUO TOSHIYUKI
            PC  C12N15/12,C07K15/12,C12N5/10,C12P21/00,(C12P21/00,C12R1:91);
            strandedness: Double;
            topology: Linear;
            hypothetical: No;
            anti-sense: No;
            *source: cell_type=pre-B cell line;
            *source: cell_line=38B9;
            *source: clone-RBP-2;
            Key             Location/Qualifiers
FEATURES             Location/Qualifiers
     source          1..1581
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     CDS
                     1..1581
                     /product='Rodent Jk rearrangement signal
                     sequence binding
                     protein'.

Qy    241 GTTGCACAGAAGTCATATATGGAAATGAAAAAAAAGGTTTTTTTTGTTTGTTTCTGTATAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTTGCACAGAAGTCATATATGGAAATGAAAAAAAAGGTTTTTTTTGTTTGTTTCTGTATAT 300

Qy    301 CTTATGGGCAGCGGATGGAGAGAAGAAAGAAACAAAATGGAACGCGATGGTTGTTCTCAA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTTATGGGCAGCGGATGGAGAGAAGAAAGAAACAAAATGGAACGCGATGGTTGTTCTCAA 360

Qy    361 CAAGAGTCTCAACCGGTGTGCATTTATTTATTGGGATAGGAAAATAGTGACACCAAGAAATGCAGCAG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAAGAGTCTCAACCGGTGTGCATTTATTTATTGGGATAGGAAAATAGTGACACCAAGAAATGCAGCAG 420

Qy    421 CTAAACTTGGAAGGAAAAGAACTATTGCACAGCCAAAACATTGTATATATATCTGACTCAGAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTAAACTTGGAAGGAAAAGAACTATTGCACAGCCAAAACATTGTATATATATCTGACTCAGAC 480

Qy    481 AAGCGAAAAGCACTTCATTTT 500
          |||||||||||||||||||||
Db    481 AAGCGAAAAGCACTTCATTTT 500
```

RESULT 4

```
LOCUS       MMJKRSB       3986 bp    mRNA       linear   ROD 29-APR-1991
DEFINITION  Mouse mRNA for J kappa RS-binding protein.
ACCESSION   X17459
VERSION     X17459.1  GI:52756
KEYWORDS    integrase; recombinase; site-specific recombination.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3986)
  AUTHORS   Matsunami,N., Hamaguchi,Y., Yamamoto,Y., Kuze,K., Kangawa,K.,
            Matsuo,H., Kawaichi,M. and Honjo,T.
  TITLE     A protein binding to the J kappa recombination sequence of
            immunoglobulin genes contains a sequence related to the integrase
            motif
  JOURNAL   Nature 342 (6252), 934-937 (1989)
  MEDLINE   90081876
FEATURES             Location/Qualifiers
     source          1..3986
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="38 B9"
                     /cell_type="pre B"
     mRNA            1..3986
                     /product="J kappa RS-binding protein"
                     /note="J kappa RS-binding protein"
                     /evidence=experimental
     CDS             1..3986
                     /codon_start=1
                     /product="J kappa RS-binding protein"
                     /db_xref="GI:52757"
                     /protein_id="CAA35501.1"
                     /db_xref="MGD:MGI:95522"
                     /db_xref="SWISS-PROT:P31266"
                     /translation="MPSGFPQSPRTSPRARPKTRITGALPMDYSEGLSAERPAHAPS
                     AGKFGERPPPKRLTREAMRNYLKERGDQTVLILHAKVAQKSYGNEKRFFCPPCVYLM
                     GSGWKKKKEQMERDGCSEQESOPCAFIGIGNSDQEMQQLNLEGKNYCTAKTLYISDSD

Qy    241 GTTGCACAGAAGTCATATGGAAATGAAAAAAAAGGTTTTTTTGCCAGCTCCTTGTGTGTAT 300
          |||||||||||||| ||||||||||||||||||||||||||||| |||||||||||
Db    241 GTTGCACAGAAGTCTTACGGAAATGAAAATGAAAAAGAATTTTTTTTTGCCCTCCTTCGTGTGTAT 300

Qy    301 CTTATGGGCAGCGGATGGAGAGAAGAAAGAAACAAAATGAACGCGATGGTTGTTCTGAA 360
          |||||||||||||||| || ||||||||||| ||||||||| || |||||||||
Db    301 CTTATGGGCAGTGGTTGGAAGAAAAAGAGAACGAGAGATGGATGTTCTGAA 360

Qy    361 CAAGAGTCTCAACCGTGTGCATTTTATTTGGGATAGGAAAATAGTGACACCAAGAAATGCAGCAG 420
          |||||||||||||||||| ||||||||||||||||||||||| ||||||||
Db    361 CAAGAGTCTCAACCCTGTCGTTTATTGGGATAGGAAAATAGTGACACCAAGAAATGCAGCAG 420

Qy    421 CTAAACTTGGAAGGAAAAGAACTATTGCACAGCCAAAACATTGTATATATATCTGACTCAGAC 480
          |||||||||||||||||||| ||||||||||||||||||| ||||
Db    421 CTGAACTTGGAAGGAAGAACTACTACACAATGTTTACAATATCTGATTCAGAC 480

Qy    481 AAGCGAAAAGCACTTCATTTT 500
          ||| || |||||||||| ||
Db    481 AAGAGAAAACACATTTCATGTT 500
```

BASE COUNT   460 a   378 c   394 g   349 t
ORIGIN

Query Match   86.6%;   Score 432.8;   DB 6;   Length 1581;
Best Local Similarity 91.6%;   Pred. No. 9.8e-114;
Matches 458;   Conservative   0;   Mismatches 42;   Indels   0;   Gaps   0;

```
Qy    1 ATCCCCTCCGGTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGGCGGGTTCCTAAAACCCGG 60
        |||||||||||||||||||||||||| | |||||||||| |||| |||||||| ||||||||
```

RESULT 5

```
LOCUS       S63463                  3991 bp    mRNA            linear   ROD 07-MAY-1993
DEFINITION  J kappa recombination signal sequence binding protein [mice, pre B
            cell line 38B9, mRNA, 3991 nt].
ACCESSION   S63463
VERSION     S63463.1  GI:238152
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3991)
  AUTHORS   Hamaguchi,Y., Matsunami,N., Yamamoto,Y., Kuze,K., Kangawa,K.,
            Matsuo,H., Kawaichi,M. and Honjo,T.
  TITLE     Cloning and characterization of a protein binding to the J kappa
            recombination signal sequence of immunoglobulin genes
  JOURNAL   Adv. Exp. Med. Biol. 292, 177-186 (1991)
  MEDLINE   92058275
  REMARK    This sequence comes from Fig.6.
            GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 63463] from the original journal article.
```

```
KRKHFMLSVKMFYGNSDDIGVFLSRIKRIKVISKPSKKKQSLKNADLCIASGTKVALFNR
LRSQVVSTRYLHVEGGNFHASSQQWGAFYIHLLDDDESEGEEFTVRDGYIHYGQTVKL
VCSVGMALPRLIIRKVDKQTLALLDADDPVSQLHKCAFYLKDTERMICLSQOERIIQF
QATPCPKEONKEMINDGASWTIISTDKAEYTFYEGMGPVLAPVTPVPVVESLQLNGGG
DVAMLELTCGONFTPNLRVWFGDVEARTMYRCGESMLCVVPDISAFREGWRWRQPVQV
PVTLVRNDGVIYSTSLITFTYTPEPGPRPHCSAAGAILRANSSQVPSNESNTNSEGNYT
NASTNSTSVTSSTATVVS"
```

```
BASE COUNT   1060 a    895 c    876 g   1155 t
ORIGIN
```

Query Match   86.6%; Score 432.8; DB 10; Length 3986;
Best Local Similarity   91.6%; Pred. No. 1.2e-113;
Matches 458; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy    1 ATCCCCTCCGGTTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGCGGTTCCTAAAACCCGG   60
        |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGCCCTCCGGTTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGCGGTTCCTAAAACCCGG   60

Qy   61 ATAACCGGACGCGCTCCCCATGGACCACACCGGAGGGGCTTGCCCGCGGGAGGAGCGCGCTGCG  120
        |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATAACCGGAGCGCGCTCCCCATGGACCACACCGGAGGGGCTTGCCCGCGGGAGGAGCGCGCTGCG  120

Qy  121 CATGCTCCATCGCCTGGGAAATTTGGTGACGGCGCCTCCACCTAAACGACTTACTAGGGAA     180
        |||||||||| | |||||||||||||||||||||||||||||| ||||||||||||||||
Db  121 CATGCTCCATCGGCGGGGAAGTTTGGTGACGGCGCCTCCACCCAAAACGACTCACTAGGGAA    180

Qy  181 GCTATGCGAAAATTATTTAAAAGAACGAGGGGGATCAAAACAGTACTTATTCTTCATGCAAAA   240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GCTATGCGAAATTATTTAAAAGAACGAGGGGGATCAAAACAGTACTTATTCTTCATGCAAA     240

Qy  241 GTTGCACAGAAGTCATATGGAAAATGAAAAATGAAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTTGTATAT  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  241 GTTGCACAGAAGTCTTACGGAAAATGAAAAATGAATTTTTTTTGCCCTCCTTGTGTTGTGTTGTAT      300
```

RESULT 6

```
LOCUS       I70264                  1500 bp    DNA             linear   PAT 03-APR-1998
```

```
FEATURES             Location/Qualifiers
     source          1..3991
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
     gene            1..3991
                     /gene="J kappa recombination signal sequence binding
                     protein"
     CDS             6..1586
                     /gene="J kappa recombination signal sequence binding
                     protein, RBP-2"
                     /note="This sequence comes from Fig.6; RBP-2"
                     /codon_start=1
                     /product="J kappa recombination signal sequence binding
                     protein"
                     /protein_id="AAB20195.1"
                     /db_xref="GI:238153"
                     /translation="MPSGFPQSPRTSPRAPRKTRITGALPMDYSEGLSAEERPAHAPS
                     AGKFGERPPKPRLTREAMRNYLKERGDQTVLILHAKVAQKSYGNERRFFCPPPCVYLM
                     GSGWKKKKEQMERDGCSSEQESOPCAFIGIGNSDQEMQQLNLBGKNYCTAKYLIYISDSD
                     KRKHFMLSVKMFYGNSDDIGVFLSRIKRIKVISKPSKKKQSLKNADLCIASGTKVALFNR
                     LRSQVVSTRYLHVEGGNFHASSQQWGAFYIHLLDDDESEGEEFTVRDGYIHYGQTVKL
                     VCSVGMALPRLIIRKVDKQTLALLDADDPVSQLHKCAFYLKDTERMICLSQOERIIQF
                     QATPCPKEONKEMINDGASWTIISTDKAEYTFYEGMGPVLAPVTPVPVVESLQLNGGG
                     DVAMLELTCGONFTPNLRVWFGDVEARTMYRCGESMLCVVPDISAFREGWRWRQPVQV
                     PVTLVRNDGVIYSTSLITFTYTPEPGPRPHCSAAGAILRANSSQVPSNESNTNSEGNYT
                     NASTNSTSVTSSTATVVS"
```

```
BASE COUNT   1062 a    895 c    877 g   1157 t
ORIGIN
```

Query Match   86.6%; Score 432.8; DB 10; Length 3991;
Best Local Similarity   91.6%; Pred. No. 1.2e-113;
Matches 458; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy    1 ATCCCCTCCGGTTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGCGGTTCCTAAAACCCGG   60
        |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6 ATGCCCTCCGGTTTTTCCTCAGTCTCCACGTACGTCCCTCAAAGCGCGGTTCCTAAAACCCGG   65

Qy   61 ATAACCGGACGCGCTCCCCATGGACCACACCGGAGGGGCTTGCCCGCGGGAGGAGCGCGCTGCG  120
        |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 ATAACCGGAGCGCGCTCCCCATGGACCACACCGGAGGGGCTTGCCCGCGGGAGGAGCGCGCTGCG  125

Qy  121 CATGCTCCATCGCCTGGGAAATTTGGTGACGGCGCCTCCACCTAAACGACTTACTAGGGAA     180
        |||||||||| | |||||||||||||||||||||||||||||| ||||||||||||||||
Db  126 CATGCTCCATCGGCGGGGAAGTTTGGTGACGGCGCCTCCACCCAAAACGACTCACTAGGGAA    185

Qy  181 GCTATGCGAAAATTATTTAAAAGAACGAGGGGGATCAAAACAGTACTTATTCTTCATGCAAAA   240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  186 GCTATGCGAAATTATTTAAAAGAACGAGGGGGATCAAAACAGTACTTATTCTTCATGCAAA     245

Qy  241 GTTGCACAGAAGTCATATGGAAAATGAAAAATGAAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTTGTATAT  300
        ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||
Db  246 GTTGCACAGAAGTCTTACGGAAAATGAAAAATGAATTTTTTTTGCCCTCCTTGTGTTGTGTTGTAT      305

Qy  301 CTTATGGGCAGCGGATGGAAGAAAAAAAAAAAGAACAAAATGAACGCGGATGGTTGTTGTTCTGAA        360
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db  306 CTTATGGGCAGTGGTTGGAAGAAAAAAAAGAGCAAATGGAACGACAGATGGAGTGTTCTGAA          365

Qy  361 CAAGAGTCTCAACCGTGTCCGCATTTATTGGGATAGGAAATAGTGACCAAGAAAATGCAGCAG     420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  366 CAAGAGTCTCAACCCTGTCCGTTTTATTGGGATAGGAAATAGAAAACAGTGACCAAGAAAATGCAGCAG  425

Qy  421 CTAAACTTGGAAGGAAAGAACTATTGCACAGCCAAAACATTGTATATATCTGACTCAGAC     480
        ||| |||||||||||||||||||||||||||| ||||||||||||| | |||||||||
Db  426 CTGAACTTGGAAGGGAAGGAAGAACTACTGCACAGCCAAAACATTGTACATATCTGATTCAGAC    485

Qy  481 AAGCGAAAGCACACTTCATTTT  500
        |||| ||||||||||||||| ||
Db  486 AAGAGAAAACACTTCATGTT   505
```

DEFINITION Sequence 1 from patent US 5679525.
ACCESSION   I70264
VERSION   I70264.1 GI:3006399
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1500)
  AUTHORS   Peterson,M.Gregory. and Henkel,T.
  TITLE   Epstein-barr virus transcription factor binding assay
  JOURNAL   Patent: US 5679525-A 1 21-OCT-1997;
FEATURES   Location/Qualifiers
     source   1..1500
                     /organism="unknown"
BASE COUNT   465 a   319 c   346 g   370 t
ORIGIN

Query Match   84.4%; Score 422; DB 6; Length 1500;
Best Local Similarity   100.0%; Pred. No. 1.3e-110;
Matches   422; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   79 ATGGACCACACGGAGGGGCTTGCCGCGCGGAGGAGCCGCCGCTGCCGCGCGGCGGCGATGCTCCATCGCCTGGG 138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACCACACGGAGGGGCTTGCCGCGCGGAGGAGCCGCCGCTGCCGCGCGGCGGCGATGCTCCATCGCCTGGG   60

Qy   139 AAATTTGGTGAGCGGGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAAATTATTTA 198
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAATTTGGTGAGCGGGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAAATTATTTA 120

Qy   199 AAAGAGCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACACAGAAGTCATAT 258
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAAGAGCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACACAGAAGTCATAT 180

Qy   259 GGAAATGAAAAAAAAAGGTTTTTTTTTTGCCCACCTCCTTGTGTATATATCTTATGGGCAGCGGATGG 318
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGAAATGAAAAAAAAGGTTTTTTTTTTGCCCACCTCCTTGTGTATATATCTTATGGGCAGCGGATGG 240

Qy   319 AAGAAAAAAAAGAACAAAAATGGAACGCGATGGTTGTTGTTCTGAACAAGAGTCTCAACCGTGT 378
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AAGAAAAAAAAGAACAAAAATGGAACGCGATGGTTGTTGTTCTGAACAAGAGTCTCAACCGTGT 300

Qy   379 GCAATTTATTGGGGATAGGGAAATAGTAGTGACCAAGAAAATGCAGCGAGCTAAACTTGGAAGGAAAG 438
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCAATTTATTGGGGATAGGGAAATAGTAGTGACCAAGAAAATGCAGCGAGCTAAACTTGGAAGGAAAG 360

Qy   439 AACTATTGCCACAGCCAAAAACATTGTATATATATCTGACTCAGACACAACAGCGAAAACACTTCATT 498
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AACTATTGCCACAGCCAAAAACATTGTATATATATCTGACTCAGACACAACAGCGAAAAGCA 420

Qy   499 TT 500
           ||
Db   421 TT 422

BASE COUNT   214 a   120 c   155 g   173 t   4 others

RESULT 7
LOCUS   AR146567   666 bp   DNA   linear   PAT 08-AUG-2001
DEFINITION Sequence 7 from patent US 6218521.
ACCESSION   AR146567
VERSION   AR146567.1 GI:15109756
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 666)
  AUTHORS   Obata,Y.
  TITLE   Isolated nucleic acid molecules associated with gastric cancer and
            methods for diagnosing and treating gastric cancer
  JOURNAL   Patent: US 6218521-A 1 17-APR-2001;
FEATURES   Location/Qualifiers
     source   1..666
                     /organism="unknown"
BASE COUNT   214 a   120 c   155 g   173 t   4 others
ORIGIN

ORIGIN

Query Match   72.5%; Score 362.6; DB 6; Length 666;
Best Local Similarity   98.9%; Pred. No. 1.3e-93;
Matches   365; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   132 GCCTGGGGAAATTTGGTGAGCGGGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA 191
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 GACAGGGGAAATTTGGTGAGCGGGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA   90

Qy   192 TTATTTAAAAGAGCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGA 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 TTATTTAAAAGAGCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGA 150

Qy   252 GTCATATGGGAAATGAAAAAAAAGGTTTTTTTTTTGCCCACCTCCTTGTGTATATCTTATGGGCAG 311
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 GTCATAATGGGAAATGAAAAAAAAGGTTTTTTTTTGCCCACCTCCTTGTGTATATCTTATGGGCAG 210

Qy   312 CGGATGGGAAGAAAAAAAAGAACAAAAATGGAACGCGATGGTTGTTGTTCTGAACAAGAGTCTCA 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 TGGATGGGAAGAAAAAAAAGAACAAAAATGGAACGCGATGGTTGTTGTTCTGAACAAGAGTCTCA 270

Qy   372 ACCGTGTGCATTTATTGGGGATAGGGAAATAGTAGTGACCAAGAAAATGCAGCGAGCTAAACTTGGA 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 ACCGTGTGCATTTATTGGGGATAGGGAAATAGTAGTGACCAAGAAAATGCAGCGAGCTAAACTTGGA 330

Qy   432 AGGAAAGAACTATTGCCACAGCCAAAAACATTGTATATATCTGACTCAGACACAACAGCGAAAAGCA 491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 AGGAAAGAACTATTGCCACAGCCAAAAACATTGTATATATCTGACTCAGACACAACAGCA 390

Qy   492 CTTCATTTT 500
           |||||||||
Db   391 CTTCATGTT 399

RESULT 8
LOCUS   AR146561   707 bp   DNA   linear   PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6218521.
ACCESSION   AR146561
VERSION   AR146561.1 GI:15109750
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 707)
  AUTHORS   Obata,Y.
  TITLE   Isolated nucleic acid molecules associated with gastric cancer and
            methods for diagnosing and treating gastric cancer
  JOURNAL   Patent: US 6218521-A 1 17-APR-2001;
FEATURES   Location/Qualifiers
     source   1..707
                     /organism="unknown"
BASE COUNT   227 a   129 c   155 g   180 t   16 others
ORIGIN

Query Match   70.9%; Score 354.4; DB 6; Length 707;
Best Local Similarity   97.0%; Pred. No. 3.1e-91;
Matches   358; Conservative   0; Mismatches   11; Indels   0; Gaps   0;

Qy   132 GCCTGGGGAAATTTGGTGAGCGGGCCGGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA 191
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17 GACAGGGGAAATTTGGTGAGCGGGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA   76

Qy   192 TTATTTAAAAGAGCGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGA 251
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77 TTATTTAAAAGACCGAGGGGATCAAACACTNCTTATTCTTCATGCAAAAGTTGCACAGA 136

Qy   252 GTCATATGGGAAATGAAATGAAAAAAAAGGTTTTTTTTTTGCCCACCTCCTTGTGTATATCTTATGGGCAG 311
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 GTCATACGGAAATGAAAATGAAANNAAGGTTTTTTTTTTGCCCACCTCCTTGTGTATATCTTATGGGCAG 196

Qy   312 CGGATGGGAAGAAAAAAAAGAACAAAAATGGAACGCGATGGTTGTTGTTCTGAACAAGAGTCTCA 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 CGGATGGGAAGAAAAAAAAGAACAAAAATGGAACGCGATGGTTGTTGTTCTGAACAAGAGTCTCA 371

RESULT 9

LOCUS   AX053603   1336 bp   DNA   linear   PAT 13-JAN-2001
DEFINITION   Sequence 369 from Patent WO0073801.
ACCESSION   AX053603
VERSION   AX053603.1 GI:12227922
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1336)
  AUTHORS   Obata,Y.
  TITLE   Breast, gastric and prostate cancer associated antigens and uses
               therefor
  JOURNAL   Patent: WO 0073801-A 369 07-DEC-2000;
               LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES   Location/Qualifiers
     source   1..1336
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   400 a   239 c   356 g   333 t   8 others
ORIGIN

Query Match   69.8%; Score 348.8; DB 6; Length 1336;
Best Local Similarity   99.2%; Pred. No. 1.5e-89;
Matches 361; Conservative   0; Mismatches   2; Indels   1; Gaps   1;

Qy   197 TGGATGGAAGAAAAAAAAAAAAGAACAAATGAACACGGATGGTTGTTCTGAACAAAAGTCTCA 256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372 ACCGTGTGCATTTATTGGGATAGGAAAATAGTGACCAGCAGCTAAACTTGGA 431

Qy   257 ACCGTGTGCATTTATTGGGATAGGAAAATAGTGACCAGCAGCTAAACTTGGA 316
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257 ACCGTGTGCATTTATTGGGATAGGAAAATAGTGACCAGCAGCTAAACTTGGA 316

Qy   317 AGGAAAGAAACTATTGCACAGCCAAAAACATTGTATATATCTGACTCAGACAAGCGAAAGCA 376
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   432 AGGAAAGAAACTATTGCACAGCCAAAAACATTGTATATATCTGACTCAGACAAGCGAAAGCA 491

Qy   377 CTTTCATTTT 385
           ||||||||||
Db   492 CTTTCATTTT 500

Qy   377 GTTGCATTTTATTGGGATAGGAAAATGAGGAAAGCACTTCA 436
           |||||||||||||||||||||||||||||||||||||||||
Db   651 GTTGCATTTTATTGGGATAGGAAAATAGTGACCAAGAAAATGCAGCAGCTAAACTTGGAAGGAA 710

RESULT 10
MMPSEUD2

LOCUS   MMPSEUD2   1085 bp   DNA   linear   ROD 30-MAR-1992
DEFINITION   M.musculus RBP-Jkappa Pseudo-2 gene.
ACCESSION   X59129
VERSION   X59129.1 GI:53807
KEYWORDS   processed pseudogene; RBP-Jkappa pseudo-2 gene.
SOURCE   house mouse.
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1085)
  AUTHORS   Kawaichi,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-APR-1991) M. Kawaichi, Dept. of Medical Chemistry,
               Kyoto University, Faculty of Medicine, Yoshide, Sa-Kyoku, Kyoto
               606, JAPAN
REFERENCE   2 (bases 1 to 1085)
  AUTHORS   Kawaichi,M., Oka,C., Shibayama,S., Koromilas,A.E., Matsunami,N.,
               Hamaguchi,Y. and Honjo,T.
  TITLE   Genomic organization of mouse J kappa recombination signal binding
               protein (RBP-J kappa) gene
  JOURNAL   J. Biol. Chem. 267 (6), 4016-4022 (1992)
  MEDLINE   92151146
COMMENT   processed type pseudogene with homology to RBP Jkappa gene.
FEATURES   Location/Qualifiers
     source   1..1085
                     /organism="Mus musculus"
                     /strain="Balb/c"
                     /db_xref="taxon:10090"
     misc_feature   198..228
                     /note="lambda gt10 genomic DNA digested with EcoR I"
     gene   229..979
                     /gene="RBP-Jkappa Pseudo-2 gene"
     exon   229..365
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="exon1"
     exon   366..404
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="exon2"
     exon   405..501
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="exon3"
     exon   502..668
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="exon4"
     exon   669..841
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="exon5"
     exon   842..979
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="exon6"
     misc_feature   1080..1085
                     /gene="RBP-Jkappa Pseudo-2 gene"
                     /note="3'end EcoR I site"
                     /evidence=experimental
BASE COUNT   376 a   229 c   227 g   253 t
ORIGIN

Query Match   66.1%; Score 330.6; DB 10; Length 1085;
Best Local Similarity   82.8%; Pred. No. 2.5e-84;
Matches 415; Conservative   0; Mismatches   79; Indels   7; Gaps   3;

Qy   137 GGAAATTTGGTGACGGCGTCCACCTCAAAACGACTTACTAGGGAAGCTATGGGAAATTATT 196
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   412 GGAAATTTGGTGACGGCGCTCCACTTAAACCGACTTACTAGGGAAGCTATGCCAAAATTATT 471

Qy   197 TAAAAGAGCGGAGGGGGATCAAACAGTACTTATATTCTTCATGCAAAAAGTTGCACAGAAGTCAT 256
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   472 TAAAAGAGCGGAGGGGGATCAAACAGTACTTATTTCTTCATGCAAAAAGTTGCACAGAAGTCAT 531

Qy   257 ATGGAAATGAAAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTATATCTTTATTGGGCAGCGGAT 316
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   532 ATGGAAATGAAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTATATCTTTATGGGCAGTGGAT 591

Qy   317 GGAAGAGAAAAAGAAAAATGAAACGGCGATGGGTTGGTTGTTCTTGAACAAGAGTCTCAAACCGT 376
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   592 GGAAG-AAAAAAAGAAAAAGGTTTTTTTGCCCACCGAGGTTTCGTTCTTGAACAGAGTCTCAACCGT 650

Qy   437 AGAACTATTGCACAGCCAAAACATTGTATATATCTGACTCAGACAAGCGAAAGCACTTCA 496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 AGAACTATTGCACAGCCAAAAACATTGTATATATCTGACTCAGACAAGCGAAAGCACTTCA 770

Qy   497 TTTT 500
           |-||
Db   771 TGTT 774

Qy   4   CCCCTCCGGTTTCCTCAGTCTCCACGTACCTCCCCTCAAAGCGCCGGGTCCTAAAACCCGGATA 63
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   232 CCCTCCAGTTTCTTCTCAGTCTCCACATATGTCCAGAGGGTGTGTTCCCAAAACCCGGATA 291

Qy   64   ACCGGAGACGCGCTCCCCCATGGACCACACGGAGGGCGTTGCCCGCGGGAGGAGCGCCGCTGCCGCAT 123
           ||| ||||| ||||||||| |||||| || ||||| ||| ||||||| ||||| ||||||

```
Db  292 ACCAGAGCTCCCCTGGGACTACTTGGAGTGCTCACCGGAAAAGGGAGAGAGCCTGCACAT 351

Qy  124 GCTCCATCGCCTGGGGAAATTTGGTGACGGCGCCTCCCACCTAAACGACTTACTAGGGAAGCT 183
        ||||||||                | |||| ||| ||| ||||| |||||  |||||||||||
Db  352 GCTCCATCACTCTGGGAAGTTTGGTGACAGCGTCACCTAAACGACTCACTAGGGAAGCT 411

Qy  184 ATGCGAAATTATTTAAAAGAGCGGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTT 243
        ||||  |||||||| ||||||||||  ||| ||||||| |||||||||||||||||||||
Db  412 ATGCAAAATTATTTAAAAGAGGCGGATCAAAACTCATTCTTCATGCATAAAGTT 471

Qy  244 GCACAGAAGTCATATGTGGAAATGAAAAAAGGTGTTTTTTGCCCACCTCCTTGTGTGTATATCTTT 303
        ||  || |  |||  |||||||| ||||| ||||||||||||||||||||  |||||| ||| ||
Db  472 TCACAGAAACTGTACAGAAATTAGAAA--GATTTTTTGCCCTCCTCCTTGTGTGGTATTTT 529

Qy  304 ATGGGCAGCGGATGGAAGAAAA---------AAAAAGAACAAAATGGAACGCGATGGGTTGTTCTGA 359
        || |||||| ||||||||||||         |||||||||||||||||||||||| |||||||||
Db  530 ATGGGCAGTGGTTGTGGAAGAAAAAGAACAAAATGGAATGAAAAATGATTGTTCTGA 589

Qy  360 ACAAGAGTCTCAACCGTGTGCATTTATTGGGATAGGAAATGTGACCAAGAAAATGCAGCA 419
        |||||||||||||||||| ||| |||||||||||||||||| |||||| |||||||||
Db  590 ACAAGAGTCTCAACCCGTGTGAGTTTTATTGGGAATAGGAAATTGTCACCAAGAAAATGCAGCA 649

Qy  420 GCTAAACTTGGAGGAAGAAGAACTATTTGCACAGCCAAAACATTGTATATAATCTGACTCAGA 479
        ||||||||||||||||||||||||  | |||||||||||||||||||||||||||||||||
Db  650 GCTGAACTTGGAGGGAAGAACTACTATACAGCCAAAACA-TGTACATACCTGATTCAGA 708

Qy  480 CAAGCCGAAAGCACACTTCATTTT 500
        |||||||||||||| |||||| ||
Db  709 CAAGAGAAAGCATTTCATGTT 729
```

```
RESULT 11
LOCUS           AX201870              373 bp     DNA        linear   PAT 30-AUG-2001
DEFINITION      Sequence 2 from Patent WO0153524.
ACCESSION       AX201870
VERSION         AX201870.1  GI:15391711
KEYWORDS
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 373)
  AUTHORS       Rees,R.C., Li,G. and Mian,S.
  TITLE         Cancer associated genes and their products
  JOURNAL       Patent: WO 0153524-A 2 26-JUL-2001;
                The Nottingham Trent University (GB)
FEATURES             Location/Qualifiers
     source          1..373
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT        128 a      59 c      93 g      93 t
ORIGIN
```

```
Qy  372 ACCGTGTGCCATTTATTGGGATAGGAAATGTGACCAAGAAAATGCAGCAGCTAAACTTGGA 431
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  270 ACCGTGTGCCATTTATTGGGATAGGAAATGTGACCAAGAAAATGCAGCAGCTAAACTTGGA 329

Qy  432 AGGAAAGAACTATTGCACAGCCAAAACATTGTATATAATCTGACT 475
        |||||||||||||||||||||||||||||||||||||||||||
Db  330 AGGAAAG-ACTATTGCACAGCCAAAACATTGTATATAATCTGACT 372
```

```
RESULT 12
LOCUS           HUMRBPJKX            1599 bp     DNA        linear   PRI 26-JUN-1995
DEFINITION      Human recombination binding protein 1 (RPB-JK) pseudogene.
ACCESSION       L34543
VERSION         L34543.1  GI:871824
KEYWORDS        nuclear protein; pseudogene; recombination binding protein 1;
                transcription factor.
SOURCE          Homo sapiens DNA.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 1599)
  AUTHORS       Zhang,M., Tang,X., Jin,C., Logeat,F., Alain,I., Kondo,S., Sun,K.
                and Yokoyama,K.
  TITLE         Genomic structure and chromosomal localization of processed
                pseudogenes for human RBP-Jk
  JOURNAL       Jpn. J. Hum. Genet. 39 (4), 393-401 (1994)
  MEDLINE       95178707
  PUBMED        7873751
FEATURES             Location/Qualifiers
     source          1..1599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
BASE COUNT        498 a     317 c     358 g     426 t
ORIGIN
```

```
Qy  137 GGAAATTTGGTGACGGCGCCTCCCACCTAAACGACTTACTAGGGAAGCTATGCGAAATTATT 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137 GGAAATTTGGTGACGTGGCTTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAATTATT 196

Qy  197 TAAAAGAGCGCAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGAAGTCAT 256
        |||| | |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  197 TAAAAGGGTAAGGGGATCAAAATAGTACTTATCCTTCATGCAAAAGTTGTACAGAAGTCAT 256

Qy  257 ATGGAAATGAAAAAAGGTGTTTTTTGCCCACCTCCTTGTGTGTATATCTTTATGGGCAGCGGAT 316
        ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257 ATGGCAATCAAAAAAATTTTTTTTTGCCCTCCCCCTTGTGTATATTCTTTATGGGCAGTGGAT 316

Qy  317 GGAAGAAAAAAAAAGAACAAAATGGAACGCGATGGGTTGTTCTGAACAAGAGTCTCAACCGT 376
        |||||||||| | ||||||||||||||||||||| ||||||||||||||||||||||| ||
Db  317 GGAAGAAAAAGAACAAAATGAAATCGATGGGTTGTTCTGAACACAGCTCTCATCCAT 376

Qy  377 GTGCATTTATTGGGATAGGAAATAGTGACCAAGAAATCCAGCAGCTAAACTTGGAAGGAA 436
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  377 GTGCATTTATTGGGATAGGAAATAGTGACCAAGAAATCCAGCAGCTAAACTTGGAAGGAA 436

Qy  437 AGAACTATTGCACAGCCAAAACATTGTATATATCATCAGACAACAAGCCAAAACCACTTCA 496
        |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  437 AGAACTATTGCACAGCCAAAACATTGTACATATCTGATTCAGACAACAAGCAAAAGCGACTTCA 496

Qy  497 TTTT 500
        ||||
Db  497 TTTT 500
```

```
RESULT 13
HUMRBPJKB
```

```
Db  210 TGGATGGAGAGAAAAAAAAAGAACGAACAGAGCGCGAGATGGTTGTTCTCGAACAAGAGTCTCA 269

Qy  312 CGGAATGGAGAGAAAAAAAAGAACGAACGAGAGTCTCA 371
        |||||||||||||||||||||||||||||||||||||
Db  210 TGGATGGAGAGAAAAAAAAAGAACGAACAGAGCGCGAGATGG 269

Qy  252 GTCATAATGGAAAATGAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTATATCTTTATGGGCAG 311
        ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   90 TTATTTAAAAGAGCGGAGGGGATCAAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGAA 149

Qy  192 TTATTTAAAAGAGCGGAGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGAA 251
        || ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   30 GACAGGGAAATTTGGTGACGGCGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA 89

Qy  132 GCCTGGGGAAATTTGGTGACGGCGCCTCCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA 191
        | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 GACAGGGAAATTTGGTGACGGCGCCTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAA 89
```

```
Matches 340; Conservative  0; Mismatches  3;   Indels  1;  Gaps  1;
Best Local Similarity  98.8%;  Pred. No.1.8e-83;
Query Match  65.4%;  Score 327.2;  DB 6;  Length 373;

Matches 337; Conservative  0; Mismatches  27;  Indels  0;  Gaps  0;
Best Local Similarity  92.6%;  Pred. No. 1.8e-81;
Query Match  64.2%;  Score 320.8;  DB 9;  Length 1599;
```

```
LOCUS       HUMRBPJKB               1121 bp    DNA     linear   PRI 21-DEC-1993
DEFINITION  Human Jk-recombination signal binding protein pseudogene 1.
ACCESSION   L07873
VERSION     L07873.1  GI:190951
KEYWORDS    Jk-recombination signal binding protein.
SOURCE      Homo sapiens adult DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Amakawa,R., Jing,W., Wzawa,K., Matsunami,N., Hamaguchi,Y.,
            Matsuda,F., Kawaichi,M. and Honjo,T.
  TITLE     Human Jk recombination signal binding protein (IGKJRB) gene:
            Comparison with its murine homologue
  JOURNAL   Genomics 17, 306-315 (1993)
  MEDLINE   94010923
COMMENT     Chromosome 9p13 or 9q13.
FEATURES             Location/Qualifiers
     source          1..1121
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
     exon            1..175
                     /germline
                     /note="homologous to exon N-2"
                     /pseudo
     exon            176..214
                     /note="homologous to exon 2"
                     /pseudo
     exon            215..310
                     /note="homologous to exon 3"
                     /pseudo
     exon            311..476
                     /note="homologous to exon 4"
                     /pseudo
     exon            477..639
                     /note="homologous to exon 5"
                     /pseudo
     exon            640..775
                     /note="homologous to exon 6"
                     /pseudo
     exon            776..888
                     /note="homologous to exon 7"
                     /pseudo
     exon            889..1029
                     /note="homologous to exon 8"
                     /pseudo
     exon            1029..1121
                     /note="homologous to exon 9"
                     /pseudo
BASE COUNT      377 a    185 c    247 g    312 t
ORIGIN
```

```
Query Match             63.8%;  Score 319.2;  DB 9;  Length 1121;
Best Local Similarity   92.3%;  Pred. No. 4.9e-81;
Matches   336;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        137 GGAAATTTGGTGAGCGGCGCCTCACCTAAACGACTTACTAGGGAAGCTATCGGAAATTATT 196
              ||||||||||| || ||| |||||||||||||||||||||||||||||| |||||||||||
Db        175 GGAAATTTGGTAAGTGGCCTCCATTTAAACGACTTACTAGGGAAGCCTATGTGAAATTATT 234

Qy        197 TAAAAGAGCGAGGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGAAGTCAT 256
              ||||||||| |||||||||||||||||||||| || |||||||| |||||||||||||||
Db        235 TAAAAGGGGTAAGGGATCAAATAGTACTTATCATGCAAAAGTTGTACAGAAGTCAT 294

Qy        257 ATGGAAATGAAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTATATCTTATGGGCAGCGGAT 316
              ||||||| |||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        295 ATGGAAATGAAAAAAGGTTTTTTTGCCCCTCCCCTTGTGTATATCTTATGGGCAGTGGAT 354

Qy        317 GGAAGAAAAAAAAGACAAAATGGAACGCGCGATTGGTTGTTGTTCTGAACAAGAGTCTCAACCGT 376
              |||||||| | |||||| |||||||| |||||||| || |||||||||||||||||||||| || |
Db        355 GGAAGAGAAAATAAACAAAATGGAAATGGGGGATGGTTGTTCTGAACAAGCCTCTCCATCCAT 414
```

```
RESULT 14
LOCUS       HUMRBPJKY               1600 bp    DNA     linear   PRI 26-JUN-1995
DEFINITION  Human recombination binding protein 1 (RBP-JK) pseudogene.
ACCESSION   L34544
VERSION     L34544.1  GI:871825
KEYWORDS    nuclear protein; pseudogene; recombination binding protein 1;
            transcription factor.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1600)
  AUTHORS   Zhang,M., Tang,X., Jin,C., Logeat,F., Alain,I., Kondo,S., Sun,K.
            and Yokoyama,K.
  TITLE     Genomic structure and chromosomal localization of processed
            pseudogenes for human RBP-Jk
  JOURNAL   Jpn. J. Hum. Genet. 39 (4), 393-401 (1994)
  MEDLINE   95178707
  PUBMED    7873751
FEATURES             Location/Qualifiers
     source          1..1600
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
BASE COUNT      503 a    321 c    359 g    417 t
ORIGIN
```

```
Query Match             63.6%;  Score 317.8;  DB 9;  Length 1600;
Best Local Similarity   93.7%;  Pred. No. 1.3e-80;
Matches   342;  Conservative    0;  Mismatches   22;  Indels    1;  Gaps    1;

Qy        137 GGAAATTTGGTGAGCGGCGCCTCCACCTAAACGACTTACTAGGGAAGCTATCGGAAATTATT 196
              ||||||||||| ||| ||||||||||||||||||||||||||||||| |||||||||||
Db        137 GGAAATTTGGTAAGTGGCTTCCACCTAAACGACTTACTAGGGAAGCTATGCGAAACTATT 196

Qy        197 TAAAAGAGCGAGGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTTGCACAGAAGTCAT 256
              ||||||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        197 TAAAAGAGTAAGGGGGATCAAATAGTACTTATTCTTCATGCAAAAGTTGCACAGAAGTCAT 256

Qy        257 ATGGAAATGAAAAAAAGGTTTTTTTGCCCACCTCCTTGTGTATATCTTATGGGCAGCGGAT 316
              ||||||||||||| ||||||||||||||| ||| |||||| |||||| ||||||||||||
Db        257 ATGGGAATGAAAAAATTTTTTTCCCTCCCCCTTGTGTATATCTTATGGGCAGCGGAT 316

Qy        317 GGAAGAAAAAAAAGACAAAATGGAACGCGCGATTGGTTGTTGTTCTGAACA-AGAGTCTCAACCG 375
              ||||||| | |||| ||||||||| |||| ||||||||||||||||||||| |||||||||| ||
Db        317 GGAGAGAAAAAAAGAACACAATGGAAATGCCGACAGTTGTTCTGAACACAGGCCTCTCAACCA 376

Qy        376 TGTGCATTTTATTTGGGATAGGAAATAGTGACCAAGAAATGCAGCACTAAACTTGGAAGGA 435
              |||||||||||||||||||||||||| |||||||||||||| ||||||||| |||||||||
Db        377 TGTGCATTTTATTTGGGATAGGAAATAGTGACCAAGAAATGCAGCACTAAACTTGGAAGGA 436

Qy        436 AAGAACTATATGCACAGACCAAAACATTGTATATATCTGACTCAGACACAAGCGAAAGCAT 495
              |||||||||||||||||||| |||||| ||||||||||||| |||||||||||||||||
Db        437 AAGAACTATTGCACAGCCAAAAACAATTGTACTATACTGATTCAGACACAAGCGAAAGCACTTC 496

Qy        496 ATTTT 500
              |||||
Db        497 ATTTT 501
```

```
RESULT 15
LOCUS       MMPSEUD1
MMPSEUD1                                       1057 bp    DNA       linear   ROD 28-SEP-1992
DEFINITION  M.musculus RBP-Jkappa Pseudo-1 gene.
ACCESSION   X59130
VERSION     X59130.1  GI:53806
KEYWORDS    pseudogene; RBP-Jkappa pseudo-1 gene.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1057)
  AUTHORS   Kawaichi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1991) M. Kawaichi, Dept. of Medical Chemistry,
            Kyoto University, Faculty of Medicine, Yoshide, Sa-Kyoku, Kyoto
            606, JAPAN
REFERENCE   2  (bases 1 to 1057)
  AUTHORS   Kawaichi,M., Oka,C., Shibayama,S., Koromilas,A.E., Matsunami,N.,
            Hamaguchi,Y. and Honjo,T.
  TITLE     Genomic organization of mouse J kappa recombination signal binding
            protein (RBP-J kappa) gene.
  JOURNAL   J. Biol. Chem. 267 (6), 4016-4022 (1992)
  MEDLINE   9156146
COMMENT     processed type pseudogene with homology to RBP JKappa gene.
FEATURES             Location/Qualifiers
     source          1..1057
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     misc_feature    198..228
                     /note="homologous to RBP-JKappa 5' flanking region"
     exon            229..365
                     /note="exon1"
                     /gene="RBP-JKappa Pseudo-1 gene"
     gene            229..1057
                     /gene="RBP-JKappa Pseudo-1 gene"
     exon            366..404
                     /note="exon2"
                     /gene="RBP-JKappa Pseudo-1 gene"
     exon            405..501
                     /note="exon3"
                     /gene="RBP-JKappa Pseudo-1 gene"
     exon            502..665
                     /note="exon4"
                     /gene="RBP-JKappa Pseudo-1 gene"
     exon            666..840
                     /note="exon5"
                     /gene="RBP-JKappa Pseudo-1 gene"
     exon            841..987
                     /note="exon6"
                     /gene="RBP-JKappa Pseudo-1 gene"
     exon            979..1057
                     /note="exon7"
                     /gene="RBP-JKappa Pseudo-1 gene"
     misc_feature    1..6
                     /note="5'end EcoR I site of the 2.4 kb fragment"
                     /clone_lib="lambda gt10 genomic DNA digested with ECOR I"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /germline
                     /clone="2.4 kb EcoR I fragment"
BASE COUNT      359 a    232 c    214 g    252 t
ORIGIN
```

```
Qy        4 CCCTCGGGTTTTCCTCAGTGTCCCACTACGTCCCCTCAAAAAGCGCGTCCTTCTTAAAACCCGGATA 63
            ||||||| |||||||||||||||| ||||||| ||| ||||||||| ||||| ||||||||||||||
Db      232 CCCTCCAGTTTTCCTCAGTGTTTCCCACATAAGTCCCAGAGGGGGTGCATCCCACCAACCCGGATA 291
```

```
Qy       64 ACCGGACGCGCTCCCCATGGGACACACGCGAGGAGGGCTTGCCCGCGCGGAGGAGGCCCGCCGCGCAT 123
            ||||| ||||||||||||||||| || |  ||||  ||||||   |||| ||  |||  ||||  |||
Db      292 TCCAGAGCTCTCCCCATAGACTACTTTGAGGGGTCGCGCGTCGCGCCCAAAAAAGGAACAGCCTGCACAT 351
```

```
Qy      124 GCTCCATCGCCTGGGGAAATTTGGTGGAGCGGGCCTTCCACCTAAACCGACTTACTAGGGAAGCT 183
            ||  ||||| |||| || | ||||| || ||| || |||| || |||||| | |||||||||| |
Db      352 TCTTCATCATCGGGGACACTTGGTGAAAGGCGTCCCACCTAAAATGACTCAACAGGGGAAGCT 411
```

```
Qy      184 ATGCGAAATTATTTAAAAGAGCGAGGGGGATCAAACAGTACTTATTCTTCATGCAAAAGTT 243
            | || ||||||||  ||||| | ||||| ||||  ||  | |||  ||||||||  || ||
Db      412 ATCAGAAATTACTTAAAGAATCAGGGAGGAATCAAAACAGTACTCATTCTTCATGCAAACATT 471
```

```
Qy      244 GCACAGAAGTCATGATGAAAATGAAAAA---AGGTTTTTTTGCCCACCTCCTTGTGTAATAT 300
            | ||| |||||| | || ||| ||||||    |||||||| |||||  |||||||| |||
Db      472 GCGCAGAAGTTGTACAGAAAATGAAAAAACTTTTTTTTTTTTGCCCTCCTCCTTGTGTTTAT 531
```

```
Qy      301 CTTATGGGGCAGCGGATGGAAGAGAAA---AAAAAGAACAAAATGGAACGCGAGTGGTTGTTTC 356
            |||||||||||||||  ||||||||||   || |||||| ||||| || | ||||| |||||
Db      532 CTTATGGGCAGTGGGTTTGAAGAAAACACAGAAAAAGAACAAAATGAACCGCGGTGATGATTGTTC 591
```

```
Qy      357 TGAACAGAGTCTCAACCGTGTGCCATTTTATTGGGATAGGAGAAATAGTAGTGACCAAGAAAATGCA 416
            | |||| |      | | ||||  |||| ||| |||   | |  |||||||| | |||||  ||||||
Db      592 TCAACA------AACCCTGTGAGTTTTATTGGAACAGGAGAAATAGTGACCAAGAAAAATGCA 643
```

```
Qy      417 GCAGCTAAAACTTGGAAGAGAAAAAACTATTTGCACACGCCAAAACATTGTATATATCTTGACTC 476
            |||||||||| |||||||| || ||||| | |||| | ||||| ||||||||||||||||||||||
Db      644 GCAGCTGAACTTGAAGGGAAGAACTACTGTACAGCCAAAAACATTGTACATACCTGATTC 703
```

```
Qy      477 AGACAACGGAAAGCACTTCATTTTT 500
            ||||||| ||||||||||||||| ||
Db      704 AGACAAGAGAAAGCATTTCATGTTT 727
```

```
Search completed: October 10, 2002, 17:55:36
Job time : 719.867 secs
```

Left margin table:
```
Query Match   63.5%;
Best Local Similarity  81.3%;
Matches   410;  Conservative   0;  Mismatches   79;  Indels   15;  Gaps   3;
          Score 317.6;  DB 10;  Length 1057;
          Pred. No. 1.4e-80;
```

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.3
            Copyright (c) 1993 - 2002 Compugen Ltd.

Run on:           October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                                                 (without alignments)
                                                 14616.225 Million cell updates/sec

Title:            US-09-489-101A-11_COPY_1_500
Perfect score:    500
Sequence:         1 tggccgggggatgggcgcc..........ccgcgggttcctgctgctgctgatt 500
Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0
Searched:         1797656 seqs, 1046326829З residues

Total number of hits satisfying chosen parameters:    3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        GenEmbl:*
                  1: gb_ba:*
                  2: gb_htg:*
                  3: gb_in:*
                  4: gb_om:*
                  5: gb_ov:*
                  6: gb_pat:*
                  7: gb_ph:*
                  8: gb_pl:*
                  9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: em_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htgo_inv:*
```

SUMMARIES

```
                        %
Result    Score     Query           Match Length    DB    ID       Description
No.                 Match Length DB
-------   -----     -------------   ------------    --    --       -----------
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

No.	Score				DB	ID	Description
1	500	100.0	2509	6	AX201590	AX201590 Sequence
2	500	100.0	2509	9	HSSOX3	X71135 H.sapiens s
3	449.2	89.8	148598	9	HSBA51C14	AL121875 Human DNA
4	243	48.6	177899	2	AC094488	AC094488 Rattus no
5	224.2	44.8	18632	10	AF434675	AF434675 Mus muscu
6	176	35.2	1692	9	AF264713	AF264713 Homo sapi
7	46.4	9.3	125020	9	AF429315	AF429315 Homo sapi
8	39.2	7.8	160434	2	AC024144	AC024144 Mus muscu
9	39.2	7.8	110000	2	LMFLCHR32_12	Continuation (13 o
10	38.8	7.8	110000	2	LMFLCHR32_13	Continuation (14 o
11	38.8	7.8	110000	2	AC004801	AC004801 Homo sapi
12	37.6	7.5	193561	9	AC011896	AC011896 Homo sapi
13	37.6	7.5	211309	2	AC015693	AC015693 Homo sapi
14	37.4	7.5	249687	2	M83196	RATMTAP
15	37.4	7.5	10120	10	RATMTAP	RATMTAP
16	36.6	7.3	249687	2	AC015693	AC015693 Homo sapi
17	36.2	7.2	170889	2	AC106344	AC106344 Rattus no
18	36	7.2	160434	2	AC024144	AC024144 Mus muscu
19	35.6	7.1	139050	8	AC090054	AC090054 Oryza sat
20	35.6	7.1	143959	2	AC090055	AC090055 Oryza sat
21	35.4	7.1	300695	2	AC079431	AC079431 Mus muscu
22	35	7.0	63405	2	AC103744	AC103744 Homo sapi
23	35	7.0	101509	2	AC027353	AC027353 Homo sapi
24	34.8	7.0	214911	2	AC010537	AC010537 Homo sapi
25	34.4	6.9	951	2	AY006450	AY006450 Drosophil
26	34.4	6.9	3225	10	MMMTAP1A4	AF182211 Mus muscu
27	34.2	6.8	637	9	AX267777	AX267777 Sequence
28	34.2	6.8	7078	9	AF257772	AF257772 Homo sapi
29	34.2	6.8	38936	9	AL358817	AL358817 Human DNA
30	34.2	6.8	75010	2	AC095677	AC095677 Rattus no
31	34.2	6.8	81521	9	AC006255	AC006255 Homo sapi
32	34.2	6.8	172421	2	AC069538	AC069538 Homo sapi
33	34.2	6.8	194140	2	AP000752	AP000752 Homo sapi
34	34.2	6.8	211383	2	AC097636	AC097636 Mus muscu
35	34.2	6.8	290452	2	AC079167	AC079167 Homo sapi
36	34	6.8	2959	9	AY030283	AY030283 Homo sapi
37	33.8	6.8	43295	2	AC024381	AC024381 Rattus no
38	33.8	6.8	172544	2	AC099122	AC099122 Homo sapi
39	33.8	6.8	185597	2	AL627237	AL627237 Mus muscu
40	33.8	6.8	190165	2	AC012344	AC012344 Homo sapi
41	33.4	6.7	60551	2	AC109261	AC109261 Mus muscu
42	33.4	6.7	1175777	2	AP003202	AP003202 Oryza sat
43	33.4	6.7	331552	2	AC046137	AC046137 Homo sapi
44	33.2	6.6	56306	9	HSJ698015	AL121923 Human DNA
45	33.2	6.6	63729	2	AC097798	AC097798 Rattus no

```
RESULT 1
AX201590
LOCUS       AX201590       2509 bp       DNA       linear    PAT 30-AUG-2001
DEFINITION  Sequence 11 from Patent WO0153349.
ACCESSION   AX201590
VERSION     AX201590.1  GI:15391439
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2509)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE     Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 11 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..2509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

ALIGNMENTS

BASE COUNT   456 a   818 c   757 g   478 t

ORIGIN

```
Qy      1 TGGGCCGGGGGGATGGGCGCGCCGGTCTGCCTTGACAGGGGTTGCAAAGTTGTTTGTTTTCTAAATTC   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGGCCGGGGGGATGGGCGCGCCGGTCTGCCTTGACAGGGGTTGCAAAGTTGTTTGTTTTCTAAATTC   60

Qy     61 CGAACGCGCCCCTCTGCCCCTCTCCCCCCCCCCCCAAATCTCGCTTGACAGGGGTTGCAAAGTTGTTTCTAAATTC  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGAACGCGCCCCTCTGCCCCTCTCCCCCCCCCCCCAAATCTCGCTTGACAGGGGTTGCAAAGTTGTTTCTAAATTC  120

Qy    121 GTCACCTCCTCAGGTTTCGTTCTTTCAAAACTTTTTGAAACCCTAATTGGTTGGTCCTCTGAG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTCACCTCCTCAGGTTTCGTTCTTTCAAAACTTTTTGAAACCCTAATTGGTTGGTCCTCTGAG  180

Qy    181 TGGGCCTCGTTGGGACTCCCGCCTCCTAAGTAACTCTTACCACGTCACTAGGCCAAAGAGGG  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGGCCTCGTTGGGACTCCCGCCTCCTAAGTAACTCTTACCACGTCACTAGGCCAAAGAGGG  240

Qy    241 GCGTGGGGTGAACGAAAAGGGCTCCCGAACTTTTTTTTTTTTCCAGCCAGCCGGACGGGGGC  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCGTGGGGTGAACGAAAAGGGCTCCCGAACTTTTTTTTTTTTCCAGCCAGCCGGACGGGGGC  300

Qy    301 TCGGTAAATGATTGGCCAGGGCGCATCACTGCGAACCTGTCAATCACGGGTCCTCCGGGTT  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCGGTAAATGATTGGCCAGGGCGCATCACTGCGAACCTGTCAATCACGGGTCCTCCGGGTT  360

Qy    361 GCGAGGGGCGGCGACCAAGCCCCAAACCCCGGGGAAATCCGAGCAGGTATATAAGGGCCCAGC  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCGAGGGGCGGCGACCAAGCCCCAAACCCCGGGGAAATCCGAGCAGGTATATAAGGGCCCAGC  420

Qy    421 TAGAGCCCAGGCAGGACTGTGAATGCGACCTGTTCGAGAGAACTCATCAGGTGCGAGAAGC  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TAGAGCCCAGGCAGGACTGTGAATGCGACCTGTTCGAGAGAACTCATCAGGTGCGAGAAGC  480

Qy    481 CCGCGGGGGTTCCTGCTGATTT  500
          ||||||||||||||||||||||
Db    481 CCGCGGGGGTTCCTGCTGATTT  500
```

RESULT 2

LOCUS   HSSOX3   2509 bp   DNA   linear   PRI 11-AUG-1994

DEFINITION   H.sapiens sox3 gene.

ACCESSION   X71135

VERSION   X71135.1   GI:468790

KEYWORDS   sox3 gene.

SOURCE   human

     ORGANISM   Homo sapiens

       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 2508)

     AUTHORS   Stevanovic,M.

     TITLE   Direct Submission

     JOURNAL   Submitted (06-APR-1993) M. Stevanovic, University of Cambridge,
       Genetics Dept, Downing Street, Cambridge, CB2 3EH, UK

REFERENCE   2 (bases 1 to 2509)

     AUTHORS   Stevanovic,M., Lovell-Badge,R., Collignon,J. and Goodfellow,P.N.

     TITLE   SOX3 is an X-linked gene related to SRY

     JOURNAL   Hum. Mol. Genet. 2 (12), 2013-2018 (1993)

     MEDLINE   94154672

FEATURES   Location/Qualifiers

     source   1..2509
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
     source   <1..>2508
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="X"

Query Match   100.0%;   Score 500;   DB 6;   Length 2509;

Best Local Similarity   100.0%;   Pred. No. 4.5e-135;

Matches 500;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

     TATA_signal   404..407
     gene   442..1773
             /gene="sox3"
     CDS   442..1773
             /gene="sox3"
             /codon_start=1
             /product=1
             /protein_id="CAA50465.1"
             /db_xref="GI:530020"
             /db_xref="SWISS-PROT:P41225"
             /translation="MRPVRENSSGARSPRVPADLARSILISLPPPDSILAHRPPSSAP
             TESQGLFTVAAPAPGAPSPPATLAHLLPAPAMYSLLETELKNPVGTPTGGPAA
             PGGAGKSSANAAGGANSGGGSGGSSGGASGGGGGTQDDRVKRPMNAFMVWSRGQRRKMALE
             NPKMHNSEISKRLGADWKLLTDAEKRPFIDEAKRLRAVHMKEYPDYKYRPRRKTKTLL
             KKDKYSLPSGLLPPGAAAAAAAAASPVGVGQVQRLDTYTHVNGWANGAYSLVQ
             EQLGYAQPPSMSSPPPPPALHRYDMAGLQYSPMMPPGAQSYMNVAAAAAASGYGGMA
             PSATAAAAAAYGQQPATAAAAAAAAAAMSLGPMGSVVKSEPSSPPPAIASHSQRACLG
             DLRDMISMYLPPGGDAADAASPLPGGRLHGVHQHYQGAGTAVNGTVPLTHI"
     polyA_signal   2493..2498

ORIGIN

```
Qy      1 TGGGCCGGGGGGATGGGCGCGCCGGTCTGCCTTGACAGGGGTTGCAAAGTTGTTTGTTTTCTAAATTC   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGGCCGGGGGGATGGGCGCGCCGGTCTGCCTTGACAGGGGTTGCAAAGTTGTTTGTTTTCTAAATTC   60

Qy     61 CGAACGCGCCCCTCTGCCCCTCTCCCCCCCCAAATCTCGCTTGCGTCGGGGGTTGCAAAGTTGTTTCTAAATTC  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGAACGCGCCCCTCTGCCCCTCTCCCCCCCCAAATCTCGCTTGCGTCGGGGGTTGCAAAGTTGTTTCTAAATTC  120

Qy    121 GTCACCTCCTCAGGTTTCGTTCTTTCAAAACTTTTTGAAACCCTAATTGGTTGGTCCTCTGAG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTCACCTCCTCAGGTTTCGTTCTTTCAAAACTTTTTGAAACCCTAATTGGTTGGTCCTCTGAG  180

Qy    181 TGGGCCTCGTTGGGACTCCCGCCTCCTAAGTAACTCTTACCACGTCACTAGGCCAAAGAGGG  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGGCCTCGTTGGGACTCCCGCCTCCTAAGTAACTCTTACCACGTCACTAGGCCAAAGAGGG  240

Qy    241 GCGTGGGGTGAACGAAAAGGGCTCCCGAACTTTTTTTTTTTTCCAGCCAGCCGGACGGGGGC  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCGTGGGGTGAACGAAAAGGGCTCCCGAACTTTTTTTTTTTTCCAGCCAGCCGGACGGGGGC  300

Qy    301 TCGGTAAATGATTGGCCAGGGCGCATCACTGCGAACCTGTCAATCACGGGTCCTCCGGGTT  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCGGTAAATGATTGGCCAGGGCGCATCACTGCGAACCTGTCAATCACGGGTCCTCCGGGTT  360

Qy    361 GCGAGGGGCGGCGACCAAGCCCCAAACCCCGGGGAAATCCGAGCAGGTATATAAGGGCCCAGC  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCGAGGGGCGGCGACCAAGCCCCAAACCCCGGGGAAATCCGAGCAGGTATATAAGGGCCCAGC  420

Qy    421 TAGAGCCCAGGCAGGACTGTGAATGCGACCTGTTCGAGAGAACTCATCAGGTGCGAGAAGC  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TAGAGCCCAGGCAGGACTGTGAATGCGACCTGTTCGAGAGAACTCATCAGGTGCGAGAAGC  480

Qy    481 CCGCGGGGGTTCCTGCTGATTT  500
          ||||||||||||||||||||||
Db    481 CCGCGGGGGTTCCTGCTGATTT  500
```

RESULT 3

LOCUS   HSBA51C14   148598 bp   DNA   linear   PRI 22-NOV-2001

DEFINITION   Human DNA sequence from clone RP11-51C14 on chromosome Xq26.2-27.3, complete sequence.

ACCESSION   AL121875

VERSION   AL121875.10   GI:17065932

KEYWORDS   HTG.

SOURCE   human.

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 148598)
AUTHORS   Wray,P.
TITLE     Direct Submission
JOURNAL   Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Nov 25, 2001 this sequence version replaced gi:715748.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome X, constructed by the Sanger Centre Chromosome X Mapping
          Group. Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          This sequence is the entire insert of clone RP11-51C14 This
          sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e. phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. RP11-51C14 is from
          the library RPCI-11.1 constructed by the group of Pieter de Jong.
          For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..148598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q26.2-27.3"
                     /clone_lib="RPCI-11.1"
                     /clone="RP11-51C14"
     misc_feature    19..145
                     /note="match: GSS: Em:AQ053564"
     misc_feature    complement(454..732)
                     /note="match: GSS: Em:AQ280322"
     repeat_region   777..4848
                     /note="L1PA14 repeat: matches 2059..6144 of consensus"
     repeat_region   6289..6316
                     /note="L1PA6 repeat: matches 5767..6141 of consensus"
     repeat_region   9906..9471
                     /note="14 copies 2 mer tg 100% conserved"
     repeat_region   13074..13273
                     /note="L1PBa repeat: matches -1546..-1347 of consensus"
     repeat_region   13456..13473
                     /note="509 copies 2 mer at 72% conserved"
     repeat_region   13361..14472
                     /note="253 copies 4 mer tata 72% conserved"
     misc_feature    13816..13875
                     /note="match: GSS: Em:AQ280322"
     repeat_region   14477..15500
                     /note="Tandem repeat. Single clone region"
     repeat_region   14481..15500
                     /note="256 copies 4 mer atat 64% conserved"
     repeat_region   15544..16027
                     /note="510 copies 2 mer tt 64% conserved"
     misc_feature    18064..18139
                     /note="match: GSS: Em:AQ122700"
     repeat_region   18667..18746
                     /note="38 copies 2 mer at 73% conserved"
     repeat_region   .18746
                     /note="30 copies 2 mer tt 81% conserved"
```

```
     repeat_region   21351..22883
                     /note="L1PA3 repeat: matches 4606..6140 of consensus"
     repeat_region   25890..25917
                     /note="7 copies 4 mer agat 100% conserved"
     repeat_region   25919..25950
                     /note="8 copies 4 mer atag 87% conserved"
     repeat_region   26785..27153
                     /note="THE1C repeat: matches 1..371 of consensus"
     repeat_region   27324..27488
                     /note="LTR41 repeat: matches 611..762 of consensus"
     repeat_region   30462..30497
                     /note="12 copies 3 mer taa 86% conserved"
     repeat_region   30963..32059
                     /note="L1MA3 repeat: matches 5205..6304 of consensus"
     repeat_region   32335..32876
                     /note="MLT2CA repeat: matches 1..503 of consensus"
     repeat_region   35423..35450
                     /note="7 copies 4 mer acac 92% conserved"
     repeat_region   36359..36540
                     /note="LTR33 repeat: matches 15..208 of consensus"
     repeat_region   43996..44019
                     /note="6 copies 4 mer agag 100% conserved"
     repeat_region   45277..45534
                     /note="L1PA8 repeat: matches 5904..6163 of consensus"
     misc_feature    48021..48473
                     /note="match: GSS: Em:AQ705461"
     repeat_region   49652..49770
                     /note="7 copies 17 mer 71% conserved"
     repeat_region   49655..49769
                     /note="57 copies 2 mer tt 71% conserved"
     repeat_region   49735..49770
                     /note="L1PA3 repeat: matches 6031..6146 of consensus"
     misc_feature    58947..59274
                     /note="match: STS: Em:HSC84H9"
     repeat_region   57027..57142
                     /note="L1M4b repeat: matches -276..-888 of consensus"
     misc_feature    55577..56705
                     /note="match: STS: Em:G04908"
     misc_feature    complement(51089..51346)
                     /note="match: GSS: Em:AQ705512"
     misc_feature    50986..51784
                     /note="match: GSS: Em:AQ1224435"
     misc_feature    complement(50539..50976)
                     /note="match: GSS: Em:AQ543128"
     misc_feature    complement(50527..50960)
                     /note="9 copies 4 mer tttg 97% conserved"
     repeat_region   49735..49770
                     /note="THE1C-internal repeat: matches 2..1563 of consensus"
     misc_feature    76815..78299
                     /note="match: GSS: Em:AQ492688"
     repeat_region   64743..64802
                     /note="20 copies 3 mer att 85% conserved"
     repeat_region   64427..64456
                     /note="10 copies 3 mer att 100% conserved"
     repeat_region   76451..76812
                     /note="THE1C repeat: matches 1..370 of consensus"
     misc_feature    71815..72103
                     /note="match: GSS: Em:AQ809267"
     repeat_region   82328..82842
                     /note="30 copies 2 mer ta 76% conserved"
     repeat_region   79736..79795
                     /note="THE1C repeat: matches 1..371 of consensus"
     repeat_region   86334..86359
                     /note="MER97c repeat: matches 1..72 of consensus"
     repeat_region   84166..84240
                     /note="match: GSS: Em:AQ809267"
     repeat_region   87839..87934
                     /note="13 copies 2 mer aa 96% conserved"
     misc_feature    complement(91096..91457)
                     /note="match: STS: Em:G10105"
     misc_feature    complement(91112..91420)
                     /note="24 copies 4 mer tata 74% conserved"
     misc_feature    78317..78683
                     /note="match: STS: Em:G09875"
```

```
repeat_region   91217..91246
                /note="10 copies 3 mer aat 96% conserved"
misc_feature    100952..101359
                /note="match: GSS: Em:AQ153393"
misc_feature    complement(101204..101774)
                /note="match: GSS: Em:AQ153393"
misc_feature    104417..104467
                /note="match: STS: Em:HS37K5S"
repeat_region   complement(105205..105600)
                /note="3 copies 17 mer 88% conserved"
misc_feature    105886..108596
                /note="match: STS: Em:G42697"
misc_feature    106152..106277
                /note="CpG island"
                /evidence=not_experimental
repeat_region   106500..106544
                /note="42 copies 3 mer ggc 69% conserved"
repeat_region   106845..106958
                /note="15 copies 3 mer ggc 88% conserved"
repeat_region   110128..110183
                /note="38 copies 3 mer gcc 60% conserved"
repeat_region   109989..110498
                /note="28 copies 2 mer gt 82% conserved"
misc_feature    110989..110498
                /note="CpG island"
misc_feature    112803..112846
                /note="8 copies 4 mer ttta 93% conserved"
misc_feature    121516..121613
                /note="4 mer ttta 93% conserved"
                /evidence=not_experimental
misc_feature    113124..113600
                /note="22 copies 2 mer tt 77% conserved"
repeat_region   116438..116688
                /note="CpG island"
                /evidence=not_experimental
misc_feature    116697..116818
                /note="match: GSS: Em:AQ681757"
misc_feature    116920..116985
                /note="match: GSS: Em:AQ681757"
repeat_region   117303..117336
                /note="22 copies 3 mer tca 89% conserved"
repeat_region   117305..117336
                /note="17 copies 2 mer tt 82% conserved"
repeat_region   121516..121613
                /note="4 mer ttta 93% conserved"
repeat_region   121939..121986
                /note="L1M4b repeat: matches -215..-111 of consensus"
repeat_region   121944..121987
                /note="24 copies 2 mer tt 81% conserved"
repeat_region   122355..122871
                /note="11 copies 4 mer tttg 84% conserved"
```

```
Query Match             89.8%;  Score 449.2;  DB 9;  Length 148599;
Best Local Similarity   96.0%;  Pred. No. 2.9e-120;
Matches  482;  Conservative   0;  Mismatches  18;  Indels   2;  Gaps   2;

Qy        1 TGGCCCGGGGGATGGGGCCCGGGGGCCCGGGGGTCGCCCGGGTTGCAAAGTTGTTTTCTAAATTC 60
            ||||| ||||||||||||||| |||| | ||||  ||| |||||||||||  ||| |||||||||
Db   107686 TGGCCGGGGGGATGGGCGGCGGCCGGTCGCCCGGGTCGCCCGGGTTGCAAAGTTGTTTTCTAAATTC 107627

Qy       61 CGAA-GCGCCCCTGCCCCCTCCCCCCAATCTGCTTGCGTCGCGGGGGTGGGGG 119
            ||| |||||||||||||||| ||||||||| ||||||||||||||||||||||
Db   107626 CGAAGCGGCGCCCTGCCCCTCCCCCAATCTGCTTGCGTCGCGGGGGTGGGGG 107567

Qy      120 GGTCACCTCCTCAGGTTTCGTTCTTCTTCAAACTTTTTTGAAACCCTAATTGGTGGCCTCTGA 179
            |||||| ||||||||||||||||||| ||||||||||||||||||||||||| |||||||||||
Db   107566 CGTCACCTCCTCAGGTTTCGTTCTTCTTCAAACTTTTTTGAAACCCTAATTGGTGGCCTCTGA 107507

Qy      180 GTGGGCCTCGTGGACTCCGCGCTCCTCAGTAACTCTTACCACGTCACTGAGCCAAGAGG 239
            |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   107506 GTGGGCCTCGTGGACTCCGCGCTCCTAAGTAACTCTTACCACGTCACTGAGCCAAGAGG 107447

Qy      240 GGCGTGGGGGTGAACGAAAAGGGCT-CCCGAACTTTTTTTTTTTTCCAGCCAGGCCGAACGGGG 298
            |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db   107446 GGCGTGGGGGTGAACGAAAAGGGCTCCCCGAACTTTTTTTTTTTTTCCAGCCAGGCCGAACGGGG 107387

Qy      299 GCTCGGTAATGATTGGCCAGGGGGGCCATCACTGCGAACCTCTCAATCACGGGGTCCTCCGGG 358
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   107386 GCTCGGTAATGATTGGCCAGGGGGGCCATCACTGCGAACCTCTCAATCACGGGGTCCTCCGGG 107327

Qy      359 TTGGGCAGGGGCGGACCAAGCCCCCAACCCCCGGGGAATCCGACGAGCAGGTATATAAGGGGCCCA 418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   107326 TTGGGCAGGGGCGGACCAAGCCCCCAACCCCCGGGGAATCCGACGAGCAGGTATATAAGGGGCCCA 107267

Qy      419 GCTAGAGCCCCAGGCAGACTGTGAATCCGACCTGTTCGGAGAGAACTCATCAGGTGCGAGAA 478
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   107266 GCTAGAGCCCCAGGCAGACTGTGAATCCGACCTGTTCGGAGAGAACTCATCAGGTGCGAGAA 107207

Qy      479 GCCCCGCGGGGTTCCTCGTCTGATTT 500
            ||||||||||||||||||||||||||
Db   107206 GCCCCGCGGGGTTCCTCGTCTGATTT 107185
```

RESULT 4
AC094488/c

```
LOCUS       AC094488            177899 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-4G21, *** SEQUENCING IN PROGRESS ***,
            61 unordered pieces.
ACCESSION   AC094488
VERSION     AC094488.2  GI:17941222
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 177899)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.R., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Kelly,C.S., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,P., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watling,K., Watlington,S., Williams,G., Williamson,A., Wleczyk,R.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177899)
  AUTHORS   Worley,K.C.
```

```
TITLE       Direct Submission
JOURNAL     Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Dec 20, 2001 this sequence version replaced gi:15624323.
COMMENT     Center: Baylor College of Medicine
            Genome Center
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------------------- Project Information
            Center project name: GATG
            Center clone name: CH230-4G21
            ----------------------- Summary Statistics
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
            Consensus quality: 151007 bases at least Q40
            Consensus quality: 158658 bases at least Q30
            Consensus quality: 164378 bases at least Q20
            Estimated Insert size: 158161; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation
          * NOTE: Estimated insert size may differ from sequence length
          * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 61 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1   6678: contig of 6678 bp in length
          *   6679   6778: gap of unknown length
          *   6779  13669: contig of 6891 bp in length
          *  13670  13769: gap of unknown length
          *  13770  21847: contig of 8078 bp in length
          *  21848  21947: gap of unknown length
          *  21948  29457: contig of 7510 bp in length
          *  29458  29557: gap of unknown length
          *  29558  34848: contig of 5291 bp in length
          *  34849  34948: gap of unknown length
          *  34949  38815: contig of 3867 bp in length
          *  38816  38915: gap of unknown length
          *  38916  43569: contig of 4654 bp in length
          *  43570  43669: gap of unknown length
          *  43670  48172: contig of 4503 bp in length
          *  48173  48272: gap of unknown length
          *  48273  53970: contig of 5698 bp in length
          *  53971  54070: gap of unknown length
          *  54071  57860: contig of 3790 bp in length
          *  57861  57960: gap of unknown length
          *  57961  60334: contig of 2374 bp in length
          *  60335  60434: gap of unknown length
          *  60435  62510: contig of 2076 bp in length
          *  62511  62610: gap of unknown length
          *  62611  67043: contig of 4433 bp in length
          *  67044  67143: gap of unknown length
          *  67144  67443: contig of 300 bp in length
          *  67444  67543: gap of unknown length
          *  67544  71960: contig of 4417 bp in length
          *  71961  72060: gap of unknown length
          *  72061  75368: contig of 3308 bp in length
          *  75369  75468: gap of unknown length
          *  75469  78454: contig of 2986 bp in length
          *  78455  78554: gap of unknown length
          *  78555  82404: contig of 3850 bp in length
          *  82405  82504: gap of unknown length
          *  82505  85147: contig of 2643 bp in length
          *  85148  85247: gap of unknown length
          *  85248  87719: contig of 2472 bp in length
          *  87720  87819: gap of unknown length
          *  87820  91037: contig of 3218 bp in length
          *  91038  91137: gap of unknown length
          *  91138  94059: contig of 2922 bp in length
          *  94060  94159: gap of unknown length
          *  94160  96333: contig of 2174 bp in length
          *  96334  96433: gap of unknown length
          *  96434  98111: contig of 1678 bp in length
          *  98112  98211: gap of unknown length
          *  98212 100568: contig of 2357 bp in length
          * 100569 100668: gap of unknown length
          * 100669 103403: contig of 2735 bp in length
          * 103404 103503: gap of unknown length
          * 103504 106108: contig of 2605 bp in length
          * 106109 106208: gap of unknown length
          * 106209 109656: contig of 3448 bp in length
          * 109657 109756: gap of unknown length
          * 109757 112398: contig of 2642 bp in length
          * 112399 112498: gap of unknown length
          * 112499 114239: contig of 1741 bp in length
          * 114240 114339: gap of unknown length
          * 114340 116264: contig of 1925 bp in length
          * 116265 116364: gap of unknown length
          * 116365 119588: contig of 3224 bp in length
          * 119589 119688: gap of unknown length
          * 119689 121165: contig of 1477 bp in length
          * 121166 121265: gap of unknown length
          * 121266 125107: contig of 3842 bp in length
          * 125108 125207: gap of unknown length
          * 125208 126572: contig of 1365 bp in length
          * 126573 126672: gap of unknown length
          * 126673 128984: contig of 2312 bp in length
          * 128985 129084: gap of unknown length
          * 129085 131037: contig of 1953 bp in length
          * 131038 131137: gap of unknown length
          * 131138 133478: contig of 2341 bp in length
          * 133479 133578: gap of unknown length
          * 133579 135913: contig of 2335 bp in length
          * 135914 136013: gap of unknown length
          * 136014 138653: contig of 2640 bp in length
          * 138654 138753: gap of unknown length
          * 138754 141443: contig of 2691 bp in length
          * 141444 141543: gap of unknown length
          * 141544 143609: contig of 2065 bp in length
          * 143610 143709: gap of unknown length
          * 143710 145996: contig of 2287 bp in length
          * 145997 146096: gap of unknown length
          * 146097 147914: contig of 1818 bp in length
          * 147915 148014: gap of unknown length
          * 148015 149725: contig of 1711 bp in length
          * 149726 149825: gap of unknown length
          * 149826 152559: contig of 2735 bp in length
          * 152560 152659: gap of unknown length
          * 152660 154607: contig of 1948 bp in length
          * 154608 154707: gap of unknown length
          * 154708 156079: contig of 1371 bp in length
          * 156080 156179: gap of unknown length
          * 156180 157778: contig of 1599 bp in length
          * 157779 157878: gap of unknown length
          * 157879 159180: contig of 1302 bp in length
          * 159181 159280: gap of unknown length
          * 159281 161105: contig of 1825 bp in length
          * 161106 161205: gap of unknown length
          * 161206 162720: contig of 1515 bp in length
          * 162721 162820: gap of unknown length
          * 162821 164433: contig of 1613 bp in length
          * 164434 164533: gap of unknown length
          * 164534 165941: contig of 1408 bp in length
          * 165942 166041: gap of unknown length
          * 166042 167143: contig of 1102 bp in length
          * 167144 167243: gap of unknown length
          * 167244 168728: contig of 1485 bp in length
          * 168729 168828: gap of unknown length
          * 168829 170391: contig of 1563 bp in length
          * 170392 170491: gap of unknown length
```

```
Query Match        48.6%;  Score 243;  DB 2;  Length 177899;
Best Local Similarity   75.1%;  Pred. No. 4.4e-60;
Matches 384;  Conservative  0;  Mismatches 115;  Indels 12;  Gaps 6;

Qy      1 TGGCCGGGGGATGGGGCGGCGGCGGGCCGGTGCCTTGACAGGGGTTGCAAAGTTGTTTTCTTAAAATTC   60
          |||||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 148905 TGGCCATGGGATGGGGCGGCTGCGTCAGTCGGGTCGGGGTCGCAAAGTGTGTCTTCGGAACCC   148846

Qy     61 CGAAGCCCCCCTCTGCCCCCCCAAATCTG------CTTGCGTCGGGGGGTGGGGGGTC   115
          |||||||||||||||||||||||||      ||||||||||| |||||||||||||
Db 148845 TCGGGCGCCTCCTGCCCCCTCCTGCGCCCCAAAATTCGCGTCGTGGCGCCCCTTCTCCCCCCCT   148786

Qy    116 GGGGGGGTCACCTCACTCAGGTTTCG--TTCTTTCAAACTTTTTGAAAACCCTTAATTGGTGC   173
          ||||||||| |||||||||||||||  ||||||||||||| |||||||||||||||||||||
Db 148785 CCTCCCTCACCTCACTCGGGTTTCTCCTTTCTTTCTTTCAAAACTTTTTGAGACCCTAATTGGTGGT   148726

Qy    174 CTCTGAGTGGGCCTTCGTTGGGACTCCCGCCCGCCTCTTAAGTAAACTCTTACCACGTCACTAGGCCA   233
          ||| |||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 148725 CTCGGAGCGGGCGTTCAGGGACTCCCGCCGCCTCCTGAATAAACTCTCATCACGTCACTCGCGACC   148666

Qy    234 AAGAGGGGGCGTG--GGGTGAACGAAAAGGGCTCCGGAACTTTTTTTTTTTTTCCAGGCCAGGCCG   291
          |||||||||||||  |||||||| || ||||||||||||||||||||||||||||||||||
Db 148665 AAGAGGGGTCTGGAGGGTGAACGAAAAGGGCTCCCGGAACTTTTTTTTTTTTTCCAAGCCAGGCCA   148606

Qy    292 AACGGGGGGGCTCGGTAATGAATTGGGCCATCACTGCGGGGTCCAATCACGGGTC   351
          |||||||||||| ||  ||||  || |||||  ||||||||||||||||  ||||
Db 148605 CA-GGGGGCGTCGGTGTTGATTGGCCATCACAGCTGGGTCGATCACGGGGTC   148547

Qy    352 CTCCGGGGTTGGGA-GGGGCGGGACCAAAGCCCACCCACCCGGGGAATCCGAGCCAGGTATATAA   410
          |||||||||||||  |||||||||||||||||||||||||| |||||||||||||||||||
Db 148546 CCACGGGGTGGTGAGGGGGCGGGACCCAACCCTCCCGGATCTGAGCCAGGTATATPA   148487

Qy    411 GGGGGCCCAGCTAG-AGCCCAGGCAGACTGTGAAATGGCACCCGTGTTCGAGAGAACTCATCAG   469
          |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db 148486 GGGGGCCCGGACGTTCCCGGGCAGGCTGCGGGAATGCGACCAGCTCGAGAGAACGCATCAG   148427

Qy    470 GTGCCGAGAGAAGCCCGCGCGGGGGGTTCCCGGCGTTCCCGCCGACTT   500
          |||  ||||||||||||||||||||||||||| |||||||||||||||
Db 148426 GTGAGAGAACCCGCGCGGGGGTTCCCGCCGACTT   148396

gene    <5336. .>6463
        /gene="Sox3"
CDS     5336. .6463
        /gene="Sox3"
        /note="HMG-box containing protein"
        /codon_start=1
        /product="transcription factor SOX3"
        /protein_id="AAL40744.1"
        /db_xref="GI:17530193"
        /translation="MYSLLETELIKNPVGGPPTPAAGTGVPAAPGAAGKSGANPAGGA
        GNGGSGGANGGGGGGGGGGGGSDDRVKRPMNAFMVWSRGQRRKMALENPRKMRSEISKR
        LGADWKILTDAEKRPFIDEAKRLRAVHMKEYPDYKYRPRRKTKLLKKDKYSLPGGLP
        PGAAAAAAAAAAAAAAASSPVGVGQRLDTYTHVNGWANGAYSLVQEOLGYAQPPSMSS
        PPPPALPQMHRYDMAGLQYSPMMPPGAQSYMNAAAAAAAASGYGMAPSAAAAAAAS
        YGQQPATPAAAAAAAAAMSLGPMGSVVKSEPSSPPPAIRSHSQRACLGDLRDMISMYL
        PPGGDAADAASPLPGGRLHGVHQHYQGAGTAVNGTVPLTHI"
BASE COUNT   5056 a  4477 c  4401 g  4697 t  1 others
ORIGIN
```

```
RESULT 5
LOCUS        AF434675       18632 bp    DNA     linear   ROD 12-DEC-2001
DEFINITION   Mus musculus transcription factor SOX3 (Sox3) gene, complete cds.
ACCESSION    AF434675
VERSION      AF434675.1  GI:17530192
KEYWORDS     .
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 18632)
  AUTHORS    Brunelli,S., Bell,D., Casey,E.S., Harland,R. and Lovell-Badge,R.
  TITLE      Expression of Sox3 throughout the developing central nervous system
             is dependent on the combined action of discrete, evolutionarily
             conserved, regulatory elements
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 18632)
  AUTHORS    Lovell-Badge,R. and Brunelli,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (15-OCT-2001) Developmental Genetics, National Institute
             for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA,
             United Kingdom
FEATURES             Location/Qualifiers
     source          1..18632
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
                     /chromosome="X"
     mRNA            <5336. .>6463
                     /gene="Sox3"
                     /product="transcription factor SOX3"
```

```
RESULT 6

Query Match        44.8%;  Score 224.2;  DB 10;  Length 18632;
Best Local Similarity   74.1%;  Pred. No. 1.4e-54;
Matches 380;  Conservative  0;  Mismatches 118;  Indels 15;  Gaps 7;

Qy      2 GGCCGGGGGGGATGGGGCGGCGGCGGGTCGCCTTGACAGGGTTGCCTTGCAAAGTTGTTTTTTCTTAAAATTCC   61
          ||||||||||| ||||||||||||||||  ||||||||||||  ||||||||||||||||| |||||
Db   4659 GGCCATGGGATGGGGCGGCTGCGTCGGTCAGCCTTAGTCCAAGTTCGCAAAGTGTGTTTCCGAAACCCG   4718

Qy     62 GAAGCGCCCCCTCTGCCCCCCCCCC------CAAATCTGCTTTGCGTCGGGGGGTGGGGGGTG   115
          |||||||||||||||||||||||      ||| ||||||||||| ||||||||||||||
Db   4719 CGGGCGCGTCCTCTGCCCCCTCCCCCAGTGTCTCGCCTTGTCGGCCGCGCCCTTCTCCCCCCT   4778

Qy    116 GGGGGGGTCACCTCACTCACTCAGGTT--TCGTTCTTTCAAACTTTTTTGAAACCCTAATTGGTGGC   173
          |||||||| |||||||||||||||||  |||||||||||| |||||||||||||||||||||||
Db   4779 CCTCCCTCACTCACTCGGGTTTCTCGCTTTCAAAACTTTTTGAGACCCTAATTGGTGGT   4838

Qy    174 CTCTGAGTGGGCCTTCGTTGGGACTCCCGCC--TCCTAAGTAAACTCTTACCACGTCACTAGGC   231
          ||| |||||| |||||||||||||||||||  ||||||||||||||||||||||||||||
Db   4839 CTCAGAGCGGGCGTTCGTTGGGACTCCCGCCGCCTCCTGAAAAGCTCTCATCACGTCACTCGAC   4898

Qy    232 CAAAGAGGGGGCGTG--GGGTGAACGAAAAGGGCTCCCGAAACTTTTTTTTTTTTTTCCAGGCCAGGC   289
          ||||||||||||||  ||||||||||||||||||||||||||||||||||||||  ||||| |||
Db   4899 CCAAGAGAGGGGGTCTGGAGGGTGAACGAAAAGGGCTCCCCGAACTTTTTTTTTTTTTCCAAGCTGGGC   4958

Qy    290 CGAACGGGGGGGCTCGGTAATGAATTGGGCCATCACTGCGGGGTCCAATCACGGG   349
          ||||||||||||  ||  ||||  || |||||  ||||||||||||||||| ||
Db   4959 C-ACAGGGGCGTCGGTGTTGATTGGGCCATCACAGCTGGGTCGATCACGGAG   5017

Qy    350 TCCTCCGGGGTTGCGA-GGGGCGGGACCAAAGCCCACCCACCCGGGGAATCCGAGCCAGGTATAT   408
          ||||||||||||||  |||||||||||||||||||||||||| |||||||||||||||||
Db   5018 GCCCACGGGGTGGTGAGGGGGCGGGACCCAACCCTCCCGGATCTGAGCCAGGTGAGGTATAT   5077

Qy    409 AAGGGGGCCCAGCTAG-AGCCCAGGCAGACTGTGAAATCGACCCGTGTTCGAGAGAACTCATC   467
          |||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db   5078 AAGGGACCGGGGCGTTCCCGGGCAAGCTGCGGGAATGCGACCAGCTCGAGAGAACGCATC   5137

Qy    468 AGGTGCGGAGAGAAGCCCGCGCGGGGGGTTCCCGCCGACTT   500
          |||| |||||||||||||||||||||||||||||||||
Db   5138 AGGTGAGAGAGAAGCCCGCGCGGGGTTCCCGCCGACTT   5170
```

```
RESULT 6
LOCUS        AF264713       1692 bp    DNA     linear   PRI 31-MAY-2000
DEFINITION   Homo sapiens sex determining region Y box 3 (SOX3) gene, complete
             cds.
ACCESSION    AF264713
VERSION      AF264713.1  GI:8118617
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1692)
```

```
AUTHORS   Gorry,M.C.; Hart,P.S., Sashi,V. and Hart,T.C.
TITLE     Clarification of the Genomic Sequence for Human SOX3
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 1692)
AUTHORS   Gorry,M.C.; Hart,P.S., Sashi,V. and Hart,T.C.
TITLE     Direct Submission
JOURNAL   Submitted (08-MAY-2000) Oral Medicine/Pathology, University of
          Pittsburgh, 3501 Terrace St, Pittsburgh, PA 15261, USA
FEATURES             Location/Qualifiers
     source          1..1692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq27.1"
     mRNA            <118..>1458
                     /gene="SOX3"
     gene            <118..>1458
                     /gene="SOX3"
     CDS             118..1458
                     /gene="SOX3"
                     /note="SRY"
                     /codon_start=1
                     /product="sex determining region Y box 3"
                     /protein_id="AAF73059.1"
                     /db_xref="GI:8118618"
                     /translation="MRPVRENSSGARSPRVPADLARSILISLPFPPDSLAHRPPSSAP
                     TESQGLFIVAAPAPGAPSPPATLAHLLPAPAMYSLLETELKNPVGTPTQOAAGTGGPAA
                     PGGAGKSANANAGANSGGSSGSSSGGGGGGGGQDDRVKRPMNAFMVWSRGQRRKMALE
                     NPKMHNSEISKRLGADWKLLTDAEKRPFIDEAKRLRAVHMKEYPDYKYRPRRKTKTLL
                     KKDKYSLPSGLLPPGAAANAASPVGVGQRLDTYTHVNGWANGAYSLVQ
                     EQLGYAQPSMSSPPPPALPPMHRYDMAGLQYSPMMPPGAQSYMNVAAAAASYG
                     GMAPSATAANAAYGQQPATAAAAAAAAMSLGPMGSVVKSEPSSPPPAISHSQRA
                     CLGDLRDMISMYLPPGDAADAASPLPGGRLHGVHQHYQGAGTAVNGTVPLTHI"
```

RESULT 7
```
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
```

```
BASE COUNT     303 a    617 c    531 g    241 t
ORIGIN

Query Match          35.2%;  Score 176;  DB 9;  Length 1692;
Best Local Similarity 100.0%;  Pred. No. 1.7e-40;
Matches 176;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    325 TCACTGCGAACCTGTCAATCACGGGGTCCTCCGGGTTGCGGGGGCGGGACCAACCCCCAAC 384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCACTGCGAACCTGTCAATCACGGGGTCCTCCGGGTTGCGGGGGCGGGACCAACCCCCAAC 60

Qy    385 CCCCGGGGAATTCCGACCGACCGAGTATATAAGGGGGGCCCCAGCCTAGACGCCCAGCCGG 444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCCGGGGAATTCCGACCGACCAGGTTATATAAGGGGGGCCCCAGCCTAGACGCCCAGCCGG 120

Qy    445 CGACCTGTTCGGAGAGAACTCATCAGGTGCGAGGAAGCCCGCGGGGGTTCCTGCTGATTT 500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGACCTGTTCGGAGAGAACTCATCAGGTGCGAGGAAGCCCGCGGGGGTTCCTGCTGATTT 176
```

```
PUBMED   11694876
REFERENCE 2 (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1..125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(3558..35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507..>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507..>36887)
                     /gene="JPH3"
     CDS             complement(<36507..36887)
                     /gene="JPH3"
                     /note="JP3"
                     /codon_start=1
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRENFDDGGSYCCGGWEDGKRAHGHGVGCTGPGKQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDCYGTEIYSDG"
```

```
BASE COUNT  29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN

Query Match           9.3%;  Score 46.4;  DB 9;  Length 125020;
Best Local Similarity  8.7%;  Pred. No. 0.011;
Matches 39;  Conservative 8;  Mismatches 175;  Indels 0;  Gaps 0;

Qy     48 GTTTTTCTAAATTCCGAAGCGCCCCCCTCTGCCCCCCTCCCCCCCCCAATCTGCTTGCGTCGGGGGT 107
          ::  : : :|||: :|      :: |: | ||||  ||  :| : :|:|   | |:| ||| |||||
Db  50591 SKHSSCHRRTWTDYRMRKYTYCCCCYTAYMBMBYHYCYDDBCWSGVWGSYGGWRGNMKMBDW 50650

Qy    108 GGGGGGTGGGGGGGGTCACCTCCTTCAGGTTTCGTTCTTTTCAAACTTTTTGAAACCCTAATT 167
          |: ::: |: : : | :::|  |:  :|: :|::  :|:|::|:|:| ||:|:: |: |:
Db  50651 BMDBSMMGKSMDGTDKKDYWCVSSMSVDVBWRWBDISWGSBSHYMRGVHTGVMRWVGY 50710

Qy    168 GGTGGGCCTCTGAGTGGGCCTCGTTGGACTCCCCGCCCCTCCTRAGTAACTCTTTACCAGTCACT 227
          |: ::: :| :|: || :: :|: | | : ||| || :||: :| ||  ::: ::||:|: |:|
Db  50711 GBCDTHVTYDGGHHGSWGMMKSCGYSDBGKCAKBGSKMCTGSMRYKYTCYGYGKTWSSKSA 50770

Qy    228 AGGCCAAAAGAGAGGGGCGTGGGGTGAACAGAAAAGGGGCTCCCGAACTTTTTTTTTTCCAGCCAG 287
          :|: ::::| |: |||| : |||: |:: || :::|||: ::|:|:| |::||::|: |:|::|
Db  50771 DBSYYYSBSHYBWGSYMCCRKBSYASSSVDVRYXKBTGMYRYVRSCSWYVKYBSMSVSAR 50830

Qy    288 GCCGAACGGGGCGGCTCGGTAATGATTGGCCAGGGGCGCATCACTGCGAACCTGTCAATCACG 347
          ::|: :| ||: :|: | || : |:||: :|||| ::|:|| |||||: |: |||: || |:|
Db  50831 KSKGHKVRBSSMKCSRRKDMSTSSMWRSMKDRKCSRCSHSHYSKSMGKWTSDVDRCYACS 50890

Qy    348 GGTCCTCCGGGTTGCGAGGGGCGGACCAAGACCCCCAAACCCCGGGGAAATCCGAGCAGGTATA 407
          ::|: :|: :: || :| |: :||:|| |:||||::|:||||| || |||:||:||:|||:|:
Db  50891 SBMKYKMHRKHYKKDSHBSWGKSHMKKGGAWRVRSMYRSMRCSVHDSSMVDKSRRMKGS 50950

Qy    408 TAAGGGGGGCCCCAGCTAGAGCCCCAGCCAGACGACTGTGAATGCGACCTGTTCGGAGAGAACTCATC 467
          ||: :|||::||| |||: |::|| |:||:|:|| | || :||:|| :|||: || |: |:|||| |
Db  50951 AMRGMKCYCYTSSMRSTKSRSMKSSWRWCSSKCYCYGYWSCTKMRSMGCSYSTGGSSMKC 51010

Qy    468 AGGTGCGAGAAGCCCGCGGGGTTCCTG 493
          :|: :|: |:||| :|: ||:||:|:|
Db  51011 MGKYCCWGGAGRSSRYCKSGSRMSMAG 51036
```

RESULT 8
LOCUS   AF429315/c
.LOCUS   AF429315   125020 bp   DNA   linear   PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION   AF429315.1 GI:17646244
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
  AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
           Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
           Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE   A repeat expansion in the gene encoding junctophilin-3 is
           associated with Huntington disease-like 2
  JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE 21583737
  PUBMED   11694876
REFERENCE 2 (bases 1 to 125020)
  AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE   Direct Submission
  JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
           Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES   Location/Qualifiers
     source   1..125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-like 2 (HDL2)"
     repeat_region   complement(35581..35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA   complement(<36507..>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene   complement(<36507..>36887)
                     /gene="JPH3"
     CDS   complement(<36507..36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGSKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGHKWVYKGEWTHGFGRYGYRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT   29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN

Query Match   7.9%; Score 39.4; DB 9; Length 125020;
Best Local Similarity 10.3%; Pred. No. 1.2;
Matches 50; Conservative 218; Mismatches 214; Indels 3; Gaps 1;

Qy   8 GGGATGGGGCGGGCGGCCGGCTGCCTTGACAGGGTTGCAAAGTGTTGTTTTCTAAATTCCGAACCG 67
            :::: :: :: ::: : :: : :: ::: : :: : ::: : : :: : :: :
Db   17790 GGYRKSAGSKSSRGGTGYYMKKGGGGKGSMSSKKKWKGGSSTSRRRGSSSAKSCSYMWGMSMC 17731

Qy   68 CCCCTCTGCCCCCCCCCCCCCCAATCTGCTTGCGTCCGGGGTGGGGGGGGGGGTCACCT 127
            :::: : : :: :: :: : :: : : : ::: : :::::::::::: ::
Db   17730 MSCMSMAKSYMMCYCYMYRMSSYMSYYKCYSCMGMSSTSYSYSCCMKMSSWGSCYKCMK 17671

Qy   128 CCTCAGGTTCGTTCTTTCAAACTTTTTGAAACCCTTAATTGGTGGCCTCTGAGTCGGCCT 187
            :::: :: :: :: :   :: :   :::: : : :::: :: ::: :: ::
Db   17670 YYSGWSSYSYSTSMGYYSYSTSCKKKYCSWSSMYKCKYSKSYRRKRSYYYWGGGKRAKKYYC 17611

RESULT 9
LOCUS   AC024144/c
.LOCUS   AC024144   160434 bp   DNA   linear   HTG 25-JAN-2002
DEFINITION Mus musculus chromosome 2 clone RP23-106A3 strain C57BL6/J, WORKING
           DRAFT SEQUENCE, 174 unordered pieces.
ACCESSION AC024144
VERSION   AC024144.9 GI:18092972
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
  ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 160434)
  AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
           Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
           Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE   High Throughput Mouse Sequencing
  JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160434)
  AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
           Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
           Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE   Direct Submission
  JOURNAL Submitted (25-FEB-2000) Department of Molecular Genetics, Albert
           Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
           Bronx, NY 10461, USA
           On Jan 9, 2002 this sequence version replaced gi:11094421.
COMMENT   Center: Harvard Partners Genome Center
           Center Code: HPGC
           Web site:
           http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous
           e.html
           Contact: gnktm@capecod.bwh.harvard.edu
           ------------Summary Statistics ------------
           Center project name: ABR
           Sequencing vector: pUC18; L08752
           Chemistry: Dye-terminator Big Dye; 100%
           *Consensus quality: 34047 at least Q20
           *Consensus quality: 30434 at least Q30
           *Consensus quality: 24749 at least Q40
           *Consensus quality: 30434 at least Q40
           **Estimated insert size: 156974 - sum-of-contigs
           **Estimated insert size: agarose-FP - N/A
           Quality coverage: 0.3 x in Q20 bases; sum-of-contigs estimation
           Quality coverage: agarose-FP - N/A
           ----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 174 contigs. The true order of the pieces

Qy   188 CGTGGAC--TCCCGCCCTCCTAAGTAACTCTTACCACGTCACTAGGCCAAAGAGGGGCGT 244
            :::: :: :: : : : : ::: : ::: : : :: : :: :: :
Db   17610 AGRRRMSYWKCCAKWMMSYCCWSYCMYYYSKSCTYKSSTCYKRGGIYWGSKTCYSAGG 17551

Qy   245 GGGGTGAACGAAAGGCGCTCCCGAACTTTTTTTTTTTTTCAGCCACGAGGCCGGCTCGG 304
            :::: : :: :::: ::: :: : : :::::::::: ::::::::::: ::::::
Db   17550 KSRSMYYCMWRSSKSKSSWSMSMARSWCMGWGAGYRNSKRSAGWGAGWRSSKGKRSTGMK 17491

Qy   305 TAATGATTTGGCCAGGGCGGCCATCACTCGGAACCTGTCAATCACCGGGTCCTCCGGGTTGCGA 364
            ::: : ::: :: : :   ::: :: :: : ::: :: : : :: :: ::::: :::
Db   17490 RACSKKTGSYGSTGRSMKKWKKKGYSKYSRGMKGKKKKTCYCMWKYYKRITSMCWWYYMKSW 17431

Qy   365 GGGGCGGACCAAAGCCCCAACCCCGGGGAATCCGAGCAGGTGATATAAGGGGCCCAGCTAGA 424
            :::::::::: ::: :: : :::: : :::: : :: : : : ::::::: :::::::
Db   17430 GYRKYRCCMKKKGCTGYRGMSSKSKSSGYKSMRGMRGSSYSTSCWKSSCWGYSWMKCMKYMS 17371

Qy   425 GCCCAGGCACGACTGTGAAATGCGACCTGTTCGAGAGACTCATCAGGTGCGAGAAGCCCGC 484
            :::::: : : : :: ::: ::::: :: ::::::: : :::::: :: ::: :::
Db   17370 YKKKRRSRSMRGSMSSKGWRGYAGRGCYSSSSMWSTRKRRSKCYSYKSYKKGRGRKMKGWGGMK 17311

Qy   485 GGGGTT 489
            :::::
Db   17310 RGSKY 17306

```
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*       1    871: contig of 871 bp in length
*     872    891: gap of unknown length
*     892   1463: contig of 572 bp in length
*    1464   1483: gap of unknown length
*    1484   2078: contig of 595 bp in length
*    2079   2098: gap of unknown length
*    2099   3015: contig of 917 bp in length
*    3016   3035: gap of unknown length
*    3036   3872: contig of 837 bp in length
*    3873   3892: gap of unknown length
*    3893   4674: contig of 782 bp in length
*    4675   4694: gap of unknown length
*    4695   5529: contig of 835 bp in length
*    5530   5549: gap of unknown length
*    5550   6364: contig of 815 bp in length
*    6365   6384: gap of unknown length
*    6385   7159: contig of 775 bp in length
*    7160   7179: gap of unknown length
*    7180   8306: contig of 1127 bp in length
*    8307   8326: gap of unknown length
*    8327   9018: contig of 692 bp in length
*    9019   9038: gap of unknown length
*    9039   9877: contig of 839 bp in length
*    9878   9897: gap of unknown length
*    9898  10581: contig of 684 bp in length
*   10582  10601: gap of unknown length
*   10602  11319: contig of 718 bp in length
*   11320  11339: gap of unknown length
*   11340  12085: contig of 746 bp in length
*   12086  12105: gap of unknown length
*   12106  12924: contig of 819 bp in length
*   12925  12944: gap of unknown length
*   12945  14045: contig of 1101 bp in length
*   14046  14065: gap of unknown length
*   14066  14782: contig of 717 bp in length
*   14783  14802: gap of unknown length
*   14803  15545: contig of 743 bp in length
*   15546  15565: gap of unknown length
*   15566  16663: contig of 1098 bp in length
*   16664  16683: gap of unknown length
*   16684  17551: contig of 868 bp in length
*   17552  17571: gap of unknown length
*   17572  18463: contig of 892 bp in length
*   18464  18483: gap of unknown length
*   18484  19273: contig of 790 bp in length
*   19274  19293: gap of unknown length
*   19294  20193: contig of 900 bp in length
*   20194  20213: gap of unknown length
*   20214  21145: contig of 932 bp in length
*   21146  21165: gap of unknown length
*   21166  22003: contig of 838 bp in length
*   22004  22023: gap of unknown length
*   22024  22900: contig of 877 bp in length
*   22901  22920: gap of unknown length
*   22921  23791: contig of 871 bp in length
*   23792  23811: gap of unknown length
*   23812  24710: contig of 899 bp in length
*   24711  24730: gap of unknown length
*   24731  25511: contig of 781 bp in length
*   25512  25531: gap of unknown length
*   25532  26220: contig of 689 bp in length
*   26221  26240: gap of unknown length
*   26241  26894: contig of 654 bp in length
*   26895  26914: gap of unknown length
*   26915  27828: contig of 914 bp in length
*   27829  27848: gap of unknown length
*   27849  28751: contig of 903 bp in length

*   28752  28771: gap of unknown length
*   28772  29504: contig of 733 bp in length
*   29505  29524: gap of unknown length
*   29525  30444: contig of 920 bp in length
*   30445  30464: gap of unknown length
*   30465  31394: contig of 930 bp in length
*   31395  31414: gap of unknown length
*   31415  32319: contig of 905 bp in length
*   32320  32339: gap of unknown length
*   32340  33189: contig of 850 bp in length
*   33190  33209: gap of unknown length
*   33210  34021: contig of 812 bp in length
*   34022  34041: gap of unknown length
*   34042  34901: contig of 860 bp in length
*   34902  34921: gap of unknown length
*   34922  35752: contig of 831 bp in length
*   35753  35772: gap of unknown length
*   35773  36850: contig of 1078 bp in length
*   36851  36870: gap of unknown length
*   36871  37853: contig of 983 bp in length
*   37854  37873: gap of unknown length
*   37874  38881: contig of 1008 bp in length
*   38882  38901: gap of unknown length
*   38902  39800: contig of 899 bp in length
*   39801  39820: gap of unknown length
*   39821  40559: contig of 739 bp in length
*   40560  40579: gap of unknown length
*   40580  41300: contig of 721 bp in length
*   41301  41320: gap of unknown length
*   41321  42139: contig of 819 bp in length
*   42140  42159: gap of unknown length
*   42160  42851: contig of 692 bp in length
*   42852  42871: gap of unknown length
*   42872  43550: contig of 679 bp in length
*   43551  43570: gap of unknown length

*   43571  44057: contig of 487 bp in length
*   44058  44077: gap of unknown length
*   44078  44699: contig of 622 bp in length
*   44700  44719: gap of unknown length
*   44720  45413: contig of 694 bp in length
*   45414  45433: gap of unknown length
*   45434  46062: contig of 629 bp in length
*   46063  46082: gap of unknown length
*   46083  47025: contig of 943 bp in length
*   47026  47045: gap of unknown length
*   47046  47928: contig of 883 bp in length
*   47929  47948: gap of unknown length
*   47949  48683: contig of 735 bp in length
*   48684  48703: gap of unknown length
*   48704  49516: contig of 813 bp in length
*   49517  49536: gap of unknown length
*   49537  50285: contig of 749 bp in length
*   50286  50305: gap of unknown length
*   50306  51121: contig of 816 bp in length
*   51122  51141: gap of unknown length
*   51142  51855: contig of 714 bp in length
*   51856  51875: gap of unknown length
*   51876  52702: contig of 827 bp in length
*   52703  52722: gap of unknown length
*   52723  53387: contig of 665 bp in length
*   53388  53407: gap of unknown length
*   53408  54126: contig of 719 bp in length
*   54127  54146: gap of unknown length
*   54147  55010: contig of 864 bp in length
*   55011  55030: gap of unknown length
*   55031  55716: contig of 686 bp in length
*   55717  55736: gap of unknown length
*   55737  56694: contig of 958 bp in length
*   56695  56714: gap of unknown length
*   56715  57342: contig of 628 bp in length
*   57343  57362: gap of unknown length
*   57363  58223: contig of 861 bp in length
*   58224  58243: gap of unknown length
```

```
          *    58244  59301: contig of 1058 bp in length
          *    59302  59321: gap of unknown length
          *    59322  60251: contig of 930 bp in length
          *    60252  60271: gap of unknown length
          *    60272  61132: contig of 861 bp in length
          *    61133  61152: gap of unknown length
          *    61153  61824: contig of 672 bp in length
          *    61825  61844: gap of unknown length

Query Match       Score 39.2;  DB 2;  Length 160434;
Best Local Similarity  58.6%;  Pred. No. 1.3;
Matches  68;  Conservative  0;  Mismatches  48;  Indels  0;  Gaps  0;

Qy      6  GGGGGATGGGGGGCGCCGGTCTGCCTTGACAGGGTTGCAAAGTTGTTTCTAAATTCCGAAG  65
           ||||||||||||  || | || |  || ||||||||||||
Db  23492  GGGGGGTGGGGGGAAAAAAAATATTTTTTTTTTTTCCCCCC  23433

Qy     66  CGCCCCTCTGCCCCCCCCCAAATCTGCTTGCGGTCGGGGGGGGGGGGTGGGGGGG  121
           ||||||  ||||||||        ||||||||
Db  23432  CCCCCCCCCTTTTTTTTTTGGGGGGGGGGGGGGGGGGGG  23377

RESULT 10
LMFLCHR32_12
WPCOMMENT
Sequence split into 28 fragments    LOCUS LMFLCHR32 Accession AL499622
Fragment Name        Begin        End
LMFLCHR32_00             1      110000
LMFLCHR32_01        100001      210000
LMFLCHR32_02        200001      310000
LMFLCHR32_03        300001      410000
LMFLCHR32_04        400001      510000
LMFLCHR32_05        500001      610000
LMFLCHR32_06        600001      710000
LMFLCHR32_07        700001      810000
LMFLCHR32_08        800001      910000
LMFLCHR32_09        900001     1010000
LMFLCHR32_10       1000001     1110000
LMFLCHR32_11       1100001     1210000
LMFLCHR32_12       1200001     1310000
LMFLCHR32_13       1300001     1410000
LMFLCHR32_14       1400001     1510000
LMFLCHR32_15       1500001     1610000
LMFLCHR32_16       1600001     1710000
LMFLCHR32_17       1700001     1810000
LMFLCHR32_18       1800001     1910000
LMFLCHR32_19       1900001     2010000
LMFLCHR32_20       2000001     2110000
LMFLCHR32_21       2100001     2210000
LMFLCHR32_22       2200001     2310000
LMFLCHR32_23       2300001     2410000
LMFLCHR32_24       2400001     2510000
LMFLCHR32_25       2500001     2610000
LMFLCHR32_26       2600001     2710000
LMFLCHR32_27       2700001     2727709

Continuation (13 of 28) of LMFLCHR32 from base 1200001 (AL499622 Leishmania major chromo

Query Match       Score 38.8;  DB 2;  Length 110000;
Best Local Similarity  51.8%;  Pred. No. 1.8;
Matches  88;  Conservative  0;  Mismatches  82;  Indels  0;  Gaps  0;

Qy     60  CCGAAGCGCCCCCTCGCCCCCTCCCCCCCAATCTGCTTGCGTCGGGGGGGGGGGGTGGGGG  119
           ||| |||||||||||||||||||||||||||||||||||||||||||
Db 100531  CAGACAAGCCTCCTACTCTCCGCCAACCCTCTCCGGGGGGGGGGGGGGGGGGGGGGGAGGAG  100590

Qy    120  GGTCACCTCCTCAGGTTTCGTTCGTTCTTCAAACTTTTTGAAACCCTAATTGGTGGCCTCTGA  179
           |||||| || |||||||| ||||||||||||||||||
Db 100591  GGGCACAATCCGCCATGCTCCTCCTCTGCAAGGAGTGTCGTATCCCTTTCTCTGCCTCTTCC  100650

Qy    180  GTGGGGCTCGTGGACTCCCGGCCCTCCTAAGTAACTCTTACCACGTCACTAG  229
           || | || ||||||||||||||||||||||||||||||||||
Db 100651  CTCATGCGCGCCGCGGCTCTCTTTCCGTCTGACTCTCTTTCCATGTCTCTCG  100700
```

```
RESULT 11
LMFLCHR32_13
WPCOMMENT
Sequence split into 28 fragments    LOCUS LMFLCHR32 Accession AL499622
Fragment Name        Begin        End
LMFLCHR32_00             1      110000
LMFLCHR32_01        100001      210000
LMFLCHR32_02        200001      310000
LMFLCHR32_03        300001      410000
LMFLCHR32_04        400001      510000
LMFLCHR32_05        500001      610000
LMFLCHR32_06        600001      710000
LMFLCHR32_07        700001      810000
LMFLCHR32_08        800001      910000
LMFLCHR32_09        900001     1010000
LMFLCHR32_10       1000001     1110000
LMFLCHR32_11       1100001     1210000
LMFLCHR32_12       1200001     1310000
LMFLCHR32_13       1300001     1410000
LMFLCHR32_14       1400001     1510000
LMFLCHR32_15       1500001     1610000
LMFLCHR32_16       1600001     1710000
LMFLCHR32_17       1700001     1810000
LMFLCHR32_18       1800001     1910000
LMFLCHR32_19       1900001     2010000
LMFLCHR32_20       2000001     2110000
LMFLCHR32_21       2100001     2210000
LMFLCHR32_22       2200001     2310000
LMFLCHR32_23       2300001     2410000
LMFLCHR32_24       2400001     2510000
LMFLCHR32_25       2500001     2610000
LMFLCHR32_26       2600001     2710000
LMFLCHR32_27       2700001     2727709

Continuation (14 of 28) of LMFLCHR32 from base 1300001 (AL499622 Leishmania major chr

Query Match       Score 38.8;  DB 2;  Length 110000;
Best Local Similarity  51.8%;  Pred. No. 1.8;
Matches  88;  Conservative  0;  Mismatches  82;  Indels  0;  Gaps  0;

Qy     60  CCGAAGCGCCCCCTCGCCCCCTCCCCCCCAATCTGCTTGCGTCGGGGGGGGGGGGTGGGGG  119
           ||| |||||||||||||||||||||||||||||||||||||||||||
Db    531  CAGACAAGCCTCCTACTCTCCGCCAACCCTCTCCGGGGGGGGGGGGGGGGGGGGGGGAGGAG  590

Qy    120  GGTCACCTCCTCAGGTTTCGTTCGTTCTTCAAACTTTTTGAAACCCTAATTGGTGGCCTCTGA  179
           |||||| || |||||||| ||||||||||||||||||
Db    591  GGGCACAATCCGCCATGCTCCTCCTCTGCAAGGAGTGTCGTATCCCTTTCTCTGCCTCTTCC  650

Qy    180  GTGGGGCTCGTGGACTCCCGGCCCTCCTAAGTAACTCTTACCACGTCACTAG  229
           || | || ||||||||||||||||||||||||||||||||||
Db    651  CTCATGCGCGCCGCGGCTCTCTTTCCGTCTGACTCTCTTTCCATGTCTCTCG  700

RESULT 12
AC004801
LOCUS       AC004801         193561 bp    DNA      linear   PRI 02-FEB-1999
DEFINITION  Homo sapiens 12q13.1 PAC RPCI1-228P16 (Roswell Park Cancer
            Institute Human PAC Library) complete sequence.
ACCESSION   AC004801
VERSION     AC004801.1  GI:4204244
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193561)
  AUTHORS   Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
            Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorell,J.H.,
            Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
            Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
            Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
            Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
            Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
```

```
TITLE     Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
            zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
JOURNAL   Unpublished
REFERENCE 2  (bases 1 to 193561)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 193561)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (30-JAN-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4  (bases 1 to 193561)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (02-FEB-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jan 30, 1999 this sequence version replaced gi:3763910.
            INFORMATION: http://gc.bcm.tmc.edu:8008/home.html or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
FEATURES             Location/Qualifiers
     source          1..193561
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12q13.1"
                     /clone="RPCI1-228P16"
     repeat_region   complement(123..408)
                     /rpt_family="AluSg"
     repeat_region   complement(805..895)
                     /rpt_family="MIR"
     gene            complement(join(1336..1461,3799..4239,4353..4461,
                     4906..4950,5702..5867,6125..6314,7789..7866,8746..8819,
                     16323..16415,19776..19839))
                     /gene="Phosphofructokinase U24183"
     repeat_region   1837..2124
                     /rpt_family="AluSx"
     repeat_region   2129..2430
                     /rpt_family="AluSg"
     repeat_region   2432..2995
                     /rpt_family="(GA)n"
```

```
     repeat_region   2996..3293
                     /rpt_family="AluY"
     repeat_region   3358..3533
                     /rpt_family="L1MD1"
     STS             5796..6186
                     /standard_name="M59723"
     repeat_region   complement(9157..9270)
                     /rpt_family="MIR"
     repeat_region   9822..9954
                     /rpt_family="FLAM_C"
     repeat_region   complement(10598..10686)
                     /rpt_family="L2"
     repeat_region   10779..10908
                     /rpt_family="FLAM_A"
     repeat_region   complement(11422..11512)
                     /rpt_family="L2"
     repeat_region   complement(11865..12121)
                     /rpt_family="L2"
     repeat_region   13424..13538
                     /rpt_family="FLAM_A"
     repeat_region   complement(13542..13842)
                     /rpt_family="AluSx"
     repeat_region   complement(15183..15241)
                     /rpt_family="AluSp"
     repeat_region   complement(18370..18662)
                     /rpt_family="MER58B"
     repeat_region   18679..18824
                     /rpt_family="MIR"
     repeat_region   complement(18098..18312)
                     /rpt_family="L2"
     repeat_region   complement(20222..20319)
                     /rpt_family="L2"
     repeat_region   20746..20904
                     /rpt_family="L1MC2"
     repeat_region   20905..21130
                     /rpt_family="MER4B"
     repeat_region   21504..22203
                     /rpt_family="L1MC2"
     repeat_region   22404..22705
                     /rpt_family="AluJb"
     repeat_region   22711..22883
                     /rpt_family="L1M4"
     repeat_region   22945..23061
                     /rpt_family="(GGAA)n"
     repeat_region   23071..23380
                     /rpt_family="L1M4"
     repeat_region   complement(23384..23505)
                     /rpt_family="MER4D"
     repeat_region   23506..23927
                     /rpt_family="LTR7"
     repeat_region   23957..24086
                     /rpt_family="HERVH"
     repeat_region   24147..24891
                     /rpt_family="HERVH"
     repeat_region   25227..26737
                     /rpt_family="HERVH"
     repeat_region   26724..27571
                     /rpt_family="HERVH"
     repeat_region   27558..27838
                     /rpt_family="HERVH"
     repeat_region   28743..29303
                     /rpt_family="L2"
     repeat_region   29790..29878
                     /rpt_family="(GGAA)n"
     repeat_region   30042..30166
                     /rpt_family="L2"
     repeat_region   complement(32453..32481)
                     /rpt_family="(CARA)n"
     repeat_region   34536..34586
                     /rpt_family="L2"
     repeat_region   complement(34879..34911)
                     /rpt_family="(GA)n"
```

```
LOCUS       AC011896              211309 bp    DNA     linear   HTG 07-JUL-2000
ACCESSION   AC011896
RESULT 13
```

```
DEFINITION  Homo sapiens chromosome 22 clone RP11-297B9, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC011896
VERSION     AC011896.5  GI:8954273
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211309)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211309)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:8569902.
            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0297B09
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 76%
            Sequencing vector: plasmid; 24%
            Chemistry: Dye-primer ET; 76% of reads
            Chemistry: Dye-terminator Big Dye; 24% of reads
            Assembly program: Phrap; version 0.990319
            Assembly program: Phrap; version 0.990319
            Consensus quality: 198087 bases at least Q40
            Consensus quality: 202017 bases at least Q30
            Consensus quality: 204572 bases at least Q20
            Insert size: 216000; agarose-fp
            Insert size: 209509; sum-of-contigs
            Quality coverage: 4.41 in Q20 bases; agarose-fp
            Quality coverage: 4.55 in Q20 bases; sum-of-contigs
            * NOTE:This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1            1997: contig of 1997 bp in length
            *     1998            2097: gap of unknown length
            *     2098            3231: contig of 1134 bp in length
            *     3232            3331: gap of unknown length
            *     3332            5357: contig of 2026 bp in length
            *     5358            5457: gap of unknown length
            *     5458            8308: contig of 2851 bp in length
            *     8309            8408: gap of unknown length
            *     8409           11342: contig of 2934 bp in length
            *    11343           11442: gap of unknown length
            *    11443           13683: contig of 2241 bp in length
            *    13684           13783: gap of unknown length
            *    13784           17777: contig of 3994 bp in length
            *    17778           17877: gap of unknown length
            *    17878           22829: contig of 4952 bp in length
            *    22830           22929: gap of unknown length
            *    22930           27690: contig of 4761 bp in length
            *    27691           27790: gap of unknown length
            *    27791           35241: contig of 7451 bp in length
            *    35242           35341: gap of unknown length
            *    35342           46194: contig of 10853 bp in length
            *    46195           46294: gap of unknown length
            *    46295           60234: contig of 13940 bp in length
            *    60235           60334: gap of unknown length
```

```
Query Match                 45.6%;  Score 37.6;  DB 9;  Length 193561;
Best Local Similarity       7.5%;   Pred. No. 3.9;
Matches   178;  Conservative   0;  Mismatches   204;  Indels   8;  Gaps   1;

Qy         73  CTGCCCCCTCCCCAAATCTGCTTGCGTCGGGGGGTGGGGGGTGGGGGGTCACCTCTCA 132
               |  |  |||| ||||   ||||
Db     133995  CTCCTCCCTGTCCCCCACTCCCAAAACGCCTGACGCGGTGCCCCCTCCGCCTGCTCC 134054

Qy        133  GGTTTCGTTCTTTTGAAACCTTAATTCGTGGCCTTCTGAGTGGGCCTTCGTGG 192
               ||| |                 |  | ||| ||  ||
Db     134055  TTTCTACCGCTTTCCCTCTCCCTCCCATGTCTTTTCCGTCTAGGGCTCTCGG 134114

Qy        193  ACTCCCCCGCCTTCCTAAGTAAACTCTTACCACGTCACTAGGCCAAAAGAGGGGGCGTGGGGGTGAA 252
               |||  |||    |||     ||        ||     ||||||||| | ||| || |
Db     134115  CCTGCGCCTCTCGCAAAACACCCCTCCACTCCGGCAGAACTCCGAGGGGAGGGGC 134174

Qy        253  CGAAAGGGCTCCCGGAACTTTTTTTTTTCAGCCAGGCCGAACGGGGGGCTCGTGATCA 312
               ||||||   ||                ||||       | | |||||||| || |
Db     134175  CGGAGGCCCACCCTTCCCCGCCTGTGGTCAGAGGGGGGCAGCCGCCGCCCGGGGTTTGGG 134234

Qy        313  GGCCAGGCGCATCACTGGGAACCTGTCAATCA-------CGGGTCCTCCGGGTTCCGA 364
               |||     |       |||||||| |    |        |   || | ||| |||
Db     134295  GGCGAGGGGCGTGACTCCCAGAGAGGGGGGCGTTCCGGCGCTTGGGCGCCGGGTCGGGGTCGGGGTTC 134354

Qy        365  GGGGCGGGGACCAAGCCCCCAACCCCGGGGAATCCGAGCAGGTATATAAGGGGCCCCCAGCTAGA 424
               |  |  | |  ||| |      | | |||     ||||
Db     134295  GGGGAAGCGTGACTCCCAGAGAGGGGGGTCCGGCGCTTGGGCGCCGGGTCGGGGTCGGGGTTC 134354

Qy        425  GCCCAGGCAGCAGACTGTGGAATGCGACCTGTTC 454
               |||||||| ||    | |   | |  || |  |
Db     134355  GCCCAGGCGGCTTCCGGGGGGCGGGGCGGGGTTC 134384
```

```
repeat_region   complement(34919..35215)
                /rpt_family="AluSg"
repeat_region   35770..35874
                /rpt_family="MIR"
repeat_region   complement(36315..36377)
                /rpt_family="AluY"
repeat_region   complement(36694..37001)
                /rpt_family="AluSx"
repeat_region   complement(37839..38140)
                /rpt_family="AluSx"
repeat_region   complement(38240..38464)
                /rpt_family="MIR"
repeat_region   complement(38490..38795)
                /rpt_family="AluSx"
repeat_region   complement(38921..39609)
                /rpt_family="AluSx"
repeat_region   complement(39799..40104)
                /rpt_family="AluJb"
repeat_region   complement(40313..40482)
                /rpt_family="AluJo"
repeat_region   complement(40484..40781)
                /rpt_family="AluJo"
repeat_region   complement(40784..40924)
                /rpt_family="AluSq"
repeat_region   41522..41619
                /rpt_family="AluJo"
repeat_region   complement(41716..41903)
                /rpt_family="MER5A"
repeat_region   complement(42195..42320)
                /rpt_family="MER20"
repeat_region   42389..42683
                /rpt_family="AluJo"
repeat_region   complement(43589..43888)
                /rpt_family="AluSq"
repeat_region   complement(44447..44746)
                /rpt_family="AluSx"
```

```
RESULT 14
LOCUS       AC015693              249687 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-67M6 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC015693
VERSION     AC015693.5  GI:9108082
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 249687)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-67M6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 249687)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6716081.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
Query Match     7.5%;  Score 37.6;  DB 2;  Length 211309;
Best Local Similarity   56.1%;  Pred. No. 3.9;
Matches  64;  Conservative   0;  Mismatches  50;  Indels   0;  Gaps   0;

Qy     46 TTGTTTTCTAAATTCCGAAGCGCCCCTCTCCCCCTCCCCCAAATCTGCTTGCGTCGGGG 105
          |||  ||||||||||||  |||  ||||||||||  ||||||  ||||| ||||||||
Db 153263 TTGNNCCTAAATTCCCTTTGCCCNCCCNNNCCCCCCCCCCCCCCCCCCCCCCCGGG 153204

Qy    106 GTGGGGGGGTGGGGGGGGTCACCTCCTTCAGGTTTCGTTCTTTCAAACTTTTTGAAA 159
          | ||||  ||||||  |||||||||||  ||||||||  ||||| |||||||| ||
Db 153203 GGGGGGCCGGGGGGGGTTTTTTTCCCAAGGTTTTGGATTTTGAAACCCCCTGGAA 153150
```

```
BASE COUNT    58157 a   48185 c   47370 g   55781 t   1816 others
ORIGIN

FEATURES             Location/Qualifiers
     source          1..211309
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="RP11-297B9"
*  60335   69268: contig of 8934 bp in length
*  69269   69368: gap of unknown length
*  69369   83528: contig of 14160 bp in length
*  83529   83628: gap of unknown length
*  83629  101926: contig of 18298 bp in length
* 101927  102026: gap of unknown length
* 102027  126944: contig of 24918 bp in length
* 126945  127044: gap of unknown length
* 127045  153683: contig of 26639 bp in length
* 153684  153783: gap of unknown length
* 153784  180011: contig of 26228 bp in length
* 180012  180111: gap of unknown length
* 180112  211309: contig of 31198 bp in length
```

```
----------
Center: Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----------
Project Information
Center project name: L1355
Center clone name: 67_M_6
----------

* NOTE: This record contains 251 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*     1    874: contig of 874 bp in length
*   875    974: gap of 100 bp
*   975   1860: contig of 886 bp in length
*  1861   1960: gap of 100 bp
*  1961   2866: contig of 906 bp in length
*  2867   2966: gap of 100 bp
*  2967   3856: contig of 890 bp in length
*  3857   3956: gap of 100 bp
*  3957   4852: contig of 896 bp in length
*  4853   4952: gap of 100 bp
*  4953   5869: contig of 917 bp in length
*  5870   5969: gap of 100 bp
*  5970   6853: contig of 884 bp in length
*  6854   6953: gap of 100 bp
*  6954   7857: contig of 904 bp in length
*  7858   7957: gap of 100 bp
*  7958   8834: contig of 877 bp in length
*  8835   8934: gap of 100 bp
*  8935   9824: contig of 890 bp in length
*  9825   9924: gap of 100 bp
*  9925  10813: contig of 889 bp in length
* 10814  10913: gap of 100 bp
* 10914  11854: contig of 941 bp in length
* 11855  11954: gap of 100 bp
* 11955  12848: contig of 894 bp in length
* 12849  12948: gap of 100 bp
* 12949  13887: contig of 939 bp in length
* 13888  13987: gap of 100 bp
* 13988  14928: contig of 941 bp in length
* 14929  15028: gap of 100 bp
* 15029  15930: contig of 902 bp in length
* 15931  16030: gap of 100 bp
* 16031  16955: contig of 925 bp in length
* 16956  17055: gap of 100 bp
* 17056  17960: contig of 905 bp in length
* 17961  18060: gap of 100 bp
* 18061  18973: contig of 913 bp in length
* 18974  19073: gap of 100 bp
* 19074  19981: contig of 908 bp in length
* 19982  20081: gap of 100 bp
* 20082  20947: contig of 866 bp in length
* 20948  21047: gap of 100 bp
* 21048  21938: contig of 891 bp in length
* 21939  22038: gap of 100 bp
* 22039  22962: contig of 924 bp in length
* 22963  23062: gap of 100 bp
* 23063  23991: contig of 929 bp in length
* 23992  24091: gap of 100 bp
* 24092  24949: contig of 858 bp in length
* 24950  25049: gap of 100 bp
* 25050  26002: contig of 953 bp in length
* 26003  26102: gap of 100 bp
* 26103  27045: contig of 943 bp in length
```

```
*   27046  27145: gap of 100 bp
*   27146  28075: contig of 930 bp in length
*   28076  28175: gap of 100 bp
*   28176  29082: contig of 907 bp in length
*   29083  29182: gap of 100 bp
*   29183  30095: contig of 913 bp in length
*   30096  30195: gap of 100 bp
*   30196  31089: contig of 894 bp in length
*   31090  31189: gap of 100 bp
*   31190  32086: contig of 897 bp in length
*   32087  32186: gap of 100 bp
*   32187  33095: contig of 909 bp in length
*   33096  33195: gap of 100 bp
*   33196  34111: contig of 916 bp in length
*   34112  34211: gap of 100 bp
*   34212  35143: contig of 932 bp in length
*   35144  35243: gap of 100 bp
*   35244  36191: contig of 948 bp in length
*   36192  36291: gap of 100 bp
*   36292  37193: contig of 902 bp in length
*   37194  37293: gap of 100 bp
*   37294  38249: contig of 956 bp in length
*   38250  38349: gap of 100 bp
*   38350  39269: contig of 920 bp in length
*   39270  39369: gap of 100 bp
*   39370  40292: contig of 923 bp in length
*   40293  40392: gap of 100 bp
*   40393  41283: contig of 891 bp in length
*   41284  41383: gap of 100 bp
*   41384  42298: contig of 915 bp in length
*   42299  42398: gap of 100 bp
*   42399  43300: contig of 902 bp in length
*   43301  43400: gap of 100 bp
*   43401  44289: contig of 889 bp in length
*   44290  44389: gap of 100 bp
*   44390  45293: contig of 904 bp in length
*   45294  45393: gap of 100 bp
*   45394  46307: contig of 914 bp in length
*   46308  46407: gap of 100 bp
*   46408  47337: contig of 930 bp in length
*   47338  47437: gap of 100 bp
*   47438  48316: contig of 879 bp in length
*   48317  48416: gap of 100 bp
*   48417  49345: contig of 929 bp in length
*   49346  49445: gap of 100 bp
*   49446  50348: contig of 903 bp in length
*   50349  50448: gap of 100 bp
*   50449  51371: contig of 923 bp in length
*   51372  51471: gap of 100 bp
*   51472  52378: contig of 907 bp in length
*   52379  52478: gap of 100 bp
*   52479  53371: contig of 893 bp in length
*   53372  53471: gap of 100 bp
*   53472  54375: contig of 904 bp in length
*   54376  54475: gap of 100 bp
*   54476  55412: contig of 937 bp in length
*   55413  55512: gap of 100 bp
*   55513  56437: contig of 925 bp in length
*   56438  56537: gap of 100 bp
*   56538  57425: contig of 888 bp in length
*   57426  57525: gap of 100 bp
*   57526  58434: contig of 909 bp in length
*   58435  58534: gap of 100 bp
*   58535  59455: contig of 921 bp in length
*   59456  59555: gap of 100 bp
*   59556  60462: contig of 907 bp in length
*   60463  60562: gap of 100 bp
*   60563  61474: contig of 912 bp in length
*   61475  61574: gap of 100 bp
*   61575  62499: contig of 925 bp in length
*   62500  62599: gap of 100 bp
*   62600  63511: contig of 912 bp in length
*   63512  63611: gap of 100 bp
```

```
*   63612  64514: contig of 903 bp in length
*   64515  64614: gap of 100 bp
*   64615  65544: contig of 930 bp in length
*   65545  65644: gap of 100 bp
*   65645  66549: contig of 905 bp in length
*   66550  66649: gap of 100 bp
*   66650  67567: contig of 918 bp in length
*   67568  67667: gap of 100 bp
*   67668  68595: contig of 928 bp in length
*   68596  68695: gap of 100 bp
*   68696  69636: contig of 941 bp in length
*   69637  69736: gap of 100 bp
*   69737  70616: contig of 880 bp in length
*   70617  70716: gap of 100 bp
*   70717  71649: contig of 933 bp in length
*   71650  71749: gap of 100 bp
*   71750  72672: contig of 923 bp in length
*   72673  72772: gap of 100 bp

Query Match     7.5%;  Score 37.6;  DB 2;  Length 249687;
Best Local Similarity   53.2%;
Matches   58;  Conservative   0;  Mismatches  51;  Indels  0;  Gaps  0;

Qy      13 GGGGGCGCCGGTCTGCCTTGACAGGGTTGCAAAGTTGTTTTTCTAAATTCCGAAGCGCCCT 72
           ||||| ||| || || ||| |||||||| ||||||||||||| || |||| || ||||||
Db  198407 GGGGCGCGGCGGTGGGNGGGGGGGGNNTNNGNTATTTTCNCNTCCCCCCCCCCCC 198466

Qy      73 CTGCCCCCTCCCCCAAATCTGCTTGCGTCGGGGGTGGGGGGGTGGGGGGG 121
           | |||| |||||||  |||| ||  ||| |||| |||||||||||||||
Db  198467 CCCCCCCCCCCCCCNNNGCCCCNGCTGGGGGGGGGGGGGGGGGGGG 198515
```

```
RESULT 15
RATMTAP/c
LOCUS       RATMTAP     10120 bp    mRNA    linear    ROD 26-FEB-1997
DEFINITION  Rattus norvegicus microtubule-associated protein 1A MAP1A (Mtap-1)
            mRNA, complete cds.
ACCESSION   M83196
VERSION     M83196.1  GI:205537
KEYWORDS    microtubule-associated protein 1A.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 10120)
  AUTHORS   Langkopf,A., Hammarback,J.A., Muller,R., Vallee,R.B. and
            Garner,C.C.
  TITLE     Microtubule-associated proteins 1A and LC2. Two proteins encoded in
            one messenger RNA
  JOURNAL   J. Biol. Chem. 267 (23), 16561-16566 (1992)
  MEDLINE   92355629
FEATURES             Location/Qualifiers
     source          1..10120
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
                     /dev_stage="adult"
     gene            1..10120
                     /gene="Mtap-1"
     CDS             455..8779
                     /gene="Mtap-1"
                     /codon_start=1
                     /product="microtubule-associated protein 1A MAP1A"
                     /protein_id="AAB48069.1"
                     /db_xref="GI:205538"
                     /translation="MDGVAAFSEYVSETVDVPSFPDLLEPPTSGGFLKLSKPCCYIFP
                     GGRDSALFAVNGFNILVDGGSDRKSCFWKLVRHLDRIDSVLLTHIGADNLPGINGLL
                     QRKVAELEEDQSOGSSYSDMVKNILISPELGVVFRNVPDKLRLPDASRKAKRSIEEAC
                     LTLQHLNRLGIQOAEPLYRVVSNTIEPLTLFHKMGVGRLDMVVLNPVRDSKEMOFLMQK
                     WAGNSKAKIGIVLANGKEAEISVPYLTSITRALVVWLPANPTEKIVRVLFPGNAPQNKI
                     LEGLEKLRIHDFLRYPVATOKDLAGGAVPANLKPSKIKHRADSKESLKAAPKTAVSKL
```

```
AKREVLEEGAKEARSELAKELAKTEKKAKEPSEKPPEKPSSERVGESSEALKAEK
RRLIKDKAGKHLKEKISKLEEKKDKEKKEIKEKRKELKKEEGRKEKKKDAKDEKRK
DIKPEVKLSKPDLKPTPEVRKTLYKAKAPGVRKVDGRAARGEKELSSEEPRIPPAQ
KGAAPPAAVSGHRELALSSPEDLIQDEEELKREERGLLAEQRDTLGIEKPLPADATEQ
GHPSAAIQVTQPSGPVLEGEHVEREKEVVPDSGPQKGSTNRGPDSGAEVEKEKTWEE
RKQREAELGPENTAARESEAEVKEDVTEKAELEEMEETHPSDEGGEETKAESPYQKH
TQEALKASPKSREALGGRDLGFQGKAPEKETASFLSSLATPAGATEHVSYIQDETIPG
YSETEQTIISDEEIHDEPEDRERPAPPRFPTSTYDLSGPEGPPEEASQADSAVPASSSK
TYGAPETELTYPPNMVAAPLAEEEHVSSAATSITECDKLSSFATSVAEDGSVASLTAPQ
TEETGKSSLLLDTVTSIPSSRTEATQGLDYVPSAGTISPTSSLEEDKGFKSPCEEDFS
VTGESEKKGETVGRGLSGEKAVGKEEKVVVTSEKLSGQYAAVPGAGHFLPPGEPALG
EVEERCLSPDDSTVKMASPPPSGPPSAAHTPFHQSPVEDKSEPRDFQEDSWGETKHSP
GVSKEDSEEQTVKPGPEEGTSEEGKGPPTRSPQAQDMPVSIAGGDTGCTIQLLPEDDK
AIVFETGEAGSNLGAGTLPGEVRTSTEEATEPQKDEVLRFTDQSLSPEDAESLSVLSV
VSPDTTKQEATPRSPCSLKEQOPHKDLWPMVSPEDTOSLSFSESSPSKETSLDISSKQ
LSPESLGTLOFGELNLGKEERGPVMKAEDDSCHLAPVSIPEPHRATVSPTDETPAGT
LPGGSFSHSALSVDRKHSPGEITGPGGHFMTSDDSLTKSPESLSSSPAMEDLAVEWGK
APGKEPELKSETROQKGQILPEKVAVVEODLIHQKOGALDEENKPGROQDKTPEQ
KGRDLDEKDTAAELDKGPEPEKDLDREDQOGRAPPAEKDRAKSEORDLQTQATE
PRDRAQERRDSEEKDKSLELRDRTPEERDRILVQEDRAPEHSIPEPTQTDRAPDRKGT
DDKEQKEEAASEEKEQVLEQKDWALGKEGETLDQEARTAEQKDEETLKEDKTOGQKSSFV
EDKTTTSKETVLDQKSAEKADSVEOQGGNAALEKTRALGLEESSPAEGSKAREQEKKYWK
EQDVVQGWRETSPTRGEPVGGOKEPVPAWEGKSPEQEVRYWRDDITLQQDAYWRELS
CDRKVWPHELDGQGARRYCEEREOSTELDEGPDEOEITPLQHTPRSPWTSDFKDFQE
PLPQKGLEVEWRWLAESPVGLPEEEDKLTRSPFEIISPASSPPEMTGQRVPSAPGOES
PVPDTESTAPMRNEPTTPSWLARIPWPVKDRPLPPAPLSPAPAPPTRAPEPHTPVPF
PGPNVEPSITPAVPPRAPISLSKDLSPPLNGSTVSCSPDRRTPSPKETGRGHWDGT
NDSDLKGAROOPEKETRSPSPHHPMPMGHSSLWPETRAYSSLSSDSHLGSSVRPSLDF
PASAFGFSSLQPAPPQLSPAEPRSAPCGSLAFSGDRALALVPCTPTRTRHDEVLEVT
KAPSLDSSLPQLPSSPGGPLLSNLPRPASPALSEGSSSEATTPVISSVAEREPPGL
PEAGESLATRDTEOTEPEQOREPTPYDERSFQYADIYEOMLLTGLGPACPTREPPLGA
RGSWPEGAERSSRPDTLLSSEQPLRPGKSSGGPPCSLSSEVEAGPQGCATDPRPHCGE
LSPSFINPPLPPSTDDSDLSTEEARLACKGKGGRRRVGRPGATGGPCPMADETPPTSASD
SGSSQSDSDVPPETEECPSITAEAALDSDEDGDFLPVDKAGGVSGTHHPRPGHDPPT
PLPDPRPSPPRPDVCMADPEGLSSESGRVERLREKGRPGRRAPGRAKPASPARRLDIR
GKRSPTPGKGPVDRTSRTVPRPRSTPSQVTSAEKDGHGSPMSKILVNGLKAGSTALGS
KGGSGPPVVDLAYIPNHCSGKTADOQDFFRRVRASYYVVSGNDPANGEPSRAVLDALL
EGKAQWGENLQVTLIPTHDTEVTREWYQQTHEQOQOLNVLVLASSSTVVMQDESFPAC
KIEF"
```

BASE COUNT   2605 a   2915 c   2777 g   1823 t
ORIGIN

Query Match   7.5%;   Score 37.4;   DB 10;   Length 10120;
Best Local Similarity   53.8%;   Pred. No. 4.7;
Matches   77;   Conservative   0;   Mismatches   66;   Indels   0;   Gaps   0;

```
Qy   64  AGCGCCCCCTCTGCCCCCTCCCCCCAATCTGCTTGGGTCGGGGGGTGGGGGGGTC  123
         |||||  |  |||||   |  |||||  | ||  |||||||||||||||||||||||
Db 1617  AGGGCTTCGGTGGACTCGCCCCTCACCCCTTCGGACTTGGAGGGTTTTTCTGGGGGGCTTC  1558

Qy  124  ACTTCCTCAGGTTTCGTTCTTTCAAAACTTTTTGAAACCCTAAATTGGTGGCCTCTGAGTGG  183
         |||||||   ||||| | |   |||||||||   ||||  ||||| | |||||| |||||
Db 1557  TCAGACGGGTTCTTTTGCTTTCTTTTCTGTTGCTGGCTAACTCCTTGGCCAGCTTCGAGCGG  1498

Qy  184  GCCTCGTGGACTCCCGCCCTCCTA  206
         |||||||  | |||||||||||| |
Db 1497  GCCTCCTTGGCCCCCTCTTCCAA  1475
```

Search completed: October 10, 2002, 18:05:58
Job time : 1337.87 secs

```
                    GenCore version 5.1.3
            Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                                           (without alignments)
                                           14616.225 Million cell updates/sec

Title:         US-09-489-101a-12_COPY_1_500
Perfect score: 500
Sequence:      1 aagcttggtgccatctattt.........ccctgccggagccaagaag 500

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      1797656 seqs, 1046326293 residues

Total number of hits satisfying chosen parameters:        3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     GenEmbl:*
               1:  gb_ba:*
               2:  gb_htg:*
               3:  gb_in:*
               4:  gb_om:*
               5:  gb_ov:*
               6:  gb_pat:*
               7:  gb_ph:*
               8:  gb_pl:*
               9:  gb_pr:*
               10: gb_ro:*
               11: gb_sts:*
               12: gb_sy:*
               13: gb_un:*
               14: gb_vi:*
               15: gb_ba:*
               16: em_fun:*
               17: em_hum:*
               18: em_in:*
               19: em_mu:*
               20: em_om:*
               21: em_or:*
               22: em_pat:*
               23: em_ph:*
               24: em_pl:*
               25: em_ro:*
               26: em_sts:*
               27: em_un:*
               28: em_vi:*
               29: em_un:*
               30: em_vi:*
               31: em_htg_hum:*
               32: em_htg_inv:*
               33: em_htg_other:*
               34: em_htgo_inv:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	DB	ID	Description
1	500	100.0	8372	6	AX201591	AX201591 Sequence
2	500	100.0	8372	6	AF107044	AF107044 Homo sapi
3	445.8	89.2	112929	9	AL159970	AL159970 Human DNA
4	58	11.6	2984	9	AK055459	AK055459 Homo sapi
5	58	11.6	180538	9	AL137061	AL137061 Human DNA
6	44	8.8	74371	9	AC005369	AC005369 Homo sapi
7	38.8	7.8	147124	9	AC022768	AC022768 Homo sapi
8	37.4	7.5	1478	9	AK025967	AK025967 Homo sapi
9	37.4	7.5	193046	2	AC023348	AC023348 Homo sapi
10	37.4	7.5	214269	9	AC167716	AC167716 Homo sapi
11	37.2	7.4	172177	9	AL359377	AL359377 Human DNA
12	37.2	7.4	233721	2	AC010159	AC010159 Homo sapi
13	37	7.4	171149	2	AL391218	AL391218 Homo sapi
14	37	7.4	204515	2	AL590128	AL590128 Homo sapi
15	36.6	7.3	146691	2	AP002402	AP002402 Homo sapi
16	36.6	7.3	151235	9	AC090936	AC090936 Homo sapi
17	36.6	7.3	159577	2	AC080060	AC080060 Homo sapi
18	36.2	7.2	76101	2	AC107595	AC107595 Rattus no
19	36.2	7.2	182945	2	AC069004	AC069004 Homo sapi
20	35.8	7.2	1359	8	ZMGS11	ZMGS11 Z.mays mRNA
21	35.8	7.2	1416	8	M2EGS1D	M2EGS1D
22	35.8	7.1	37426	2	AC079357	AC079357 Oryza sat
23	35.8	7.1	37200	1	SCE63	SCE63 Streptomy
24	35.4	7.1	194142	2	AC092634	AC092634 Homo sapi
25	35.4	7.1	209919	9	AC008397	AC008397 Homo sapi
26	35.2	7.0	160998	2	AL359192	AL359192 Homo sapi
27	35.2	7.0	166370	2	AC106804	AC106804 Homo sapi
28	35.2	7.0	206762	2	AC012184	AC012184 Homo sapi
29	35	7.0	125020	9	AF429315	AF429315 Homo sapi
30	35	7.0	213943	2	AC015465	AC015465 Homo sapi
31	34.8	7.0	66707	2	AC090721	AC090721 Homo sapi
32	34.8	7.0	102319	9	AL159153	AL159153 Human DNA
33	34.8	7.0	153937	9	AL451075	AL451075 Human DNA
34	34.6	6.9	1315	10	RATIGFBP	RATIGFBP Rat insulin
35	34.6	6.9	1355	10	BC012724	BC012724 Mus muscu
36	34.6	6.9	2433	10	MUSIGFBP01	L05436 Mouse insul
37	34.6	6.9	10964	8	SPAC521	AL133481 S.pombe c
38	34.6	6.9	100231	8	ATF22JT2	AL391734 Arabidops
39	34.6	6.9	173071	9	AC009477	AC009477 Homo sapi
40	34.6	6.9	177241	2	AC029960	AC029960 Rattus no
41	34.6	6.9	199508	2	AC095555	AC095555 Rattus no
42	34.6	6.9	207241	2	AC098711	AC098711 Mus muscu
43	34.4	6.9	127913	9	AL590727	AL590727 Human DNA
44	34.4	6.9	159057	2	AC023995	AC023995 Homo sapi
45	34.4	6.9	174723	9	AC018676	AC018676 Homo sapi

ALIGNMENTS

```
RESULT 1
AX201591
LOCUS       AX201591     8372 bp    DNA    linear   PAT 30-AUG-2001
DEFINITION  Sequence 12 from Patent WO0153349.
ACCESSION   AX201591
VERSION     AX201591.1  GI:15391440
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 8372)
AUTHORS     Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
TITLE       Small cell lung cancer associated antigens and uses therefor
JOURNAL     Patent: WO 0153349-A 12 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..8372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

RESULT 2

BASE COUNT   1996 a   2143 c   2104 g   2124 t   5 others
ORIGIN

Query Match   Best Local Similarity   100.0%;   Score 500;   DB 6;   Length 8372;
Matches 500;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
Pred. No. 1.8e-144;

```
Qy      1 AAGCTTGGTGTGCCATCTATTTTGGACTATGCCTGCATACAGCTTATGGGAACATTTGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTGGTGTGCCATCTATTTTGGACTATGCCTGCATACAGCTTATGGGAACATTTGTC 60

Qy     61 AGGCAAAAGTATAATAATGGCAAACTCTACGCCTTTTATTTTAAATTAGATTGGTGTGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGCAAAAGTATAATAATGGCAAACTCTACGCCTTTTATTTTAAATTAGATTGGTGTGAT 120

Qy    121 TTGATGCTGACGGGAGTGAGAGTAATGGCCTTATCCTGCTGCAGGCCTGTGCTGAGGATGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGATGCTGACGGGAGTGAGAGTAATGGCCTTATCCTGCTGCAGGCCTGTGCTGAGGATGG 180

Qy    181 CCTGGTCTCTGCCACCCTCCTCGAGTAGCCATTTTGCATGTGTAACAGGGGTCTCCCCTCTGGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTGGTCTCTGCCACCCTCCTCGAGTAGCCATTTTGCATGTGTAACAGGGGTCTCCCCTCTGGG 240

Qy    241 GCACAACAACAAAAGAGAAGAGTTGCTAAGGACAAGACAGGTGCGGAAATGCATCTCCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCACAACAACAAAAGAGAAGAGTTGCTAAGGACAAGACAGGTGCGGAAATGCATCTCCAT 300

Qy    301 TGGAACAGCCCTGGGCTTACTCCAATGGCTGAGAGGTGGCTATGGCCAGTCCTCCCCAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGAACAGCCCTGGGCTTACTCCAATGGCTGAGAGGTGGCTATGGCCAGTCCTCCCCAGA 360

Qy    361 GCTCTGCAGCTGCACTTTGGGGGGTGGAGGTGTCCTGCTTGTCGTCGTGTGTCGTGATAACGGCCGT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTCTGCAGCTGCACTTTGGGGGGTGGAGGTGTCCTGCTTGTCGTCGTGTGTCGTGATAACGGCCGT 420

Qy    421 GAAAGCCAGCCAACTGCTGCCCAAAATCACCCAGCCGATTGGGGGTTTCCCATCGGCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAAGCCAGCCAACTGCTGCCCAAAATCACCCAGCCGATTGGGGGTTTCCCATCGGCGCA 480

Qy    481 CCCTGCCCCGGGAGCCAAGAAG 500
          ||||||||||||||||||||||
Db    481 CCCTGCCCCGGGAGCCAAGAAG 500
```

RESULT 2

LOCUS   AF107044   8372 bp   DNA   linear   PRI 13-DEC-1998
DEFINITION   Homo sapiens clone pCL4 DNA-binding protein SOX21 (SOX21) gene, complete cds.
ACCESSION   AF107044
VERSION   AF107044.1   GI:4008102
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 8372)
  AUTHORS   Malas,S., Duthie,S. and Episkopou,V.
  TITLE   The cloning and chromosomal localization of human SOX14 and SOX21;
            two members of the SOX gene family related to SOX1, SOX2 and SOX3
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 8372)
  AUTHORS   Malas,S., Duthie,S. and Episkopou,V.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-NOV-1998) Clinical Sciences Centre, Medical Research
            Council, Du Cane Rd, London W12 ONN, UK
FEATURES   Location/Qualifiers
     source   1..8372
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="13"
                   /map="13q32-q33"
                   /clone="pCL4"

     mRNA   <1167..>1997
                   /gene="SOX21"
                   /product="DNA-binding protein SOX21"
     gene   <1167..>1997
                   /gene="SOX21"
     CDS   1167..1997
                   /gene="SOX21"
                   /note="Sry related HMG-box containing gene 21"
                   /codon_start=1
                   /product="DNA-binding protein SOX21"
                   /protein_id="AAC95381.1"
                   /db_xref="GI:4008103"
                   /translation="MSKPVDHVKRPMNAFMVWSRAQRKMAQENPKMHNSEISKRLGA
                   EWKLLTESERPFIDEAKRLRAMHMKEHPDYKYRPRRKPKTLLKKDKYFVPVYGLGG
                   VADAEHPALKAGAGLHAGAGGLVPESLLANPEKAAAAAAAVRVFPQSAAAAAA
                   AAAAAAGSPYSLLDLGSKMAEISSSSSGLPYASSLGYTPAGAFHGAAAAAAAA
                   AGGHHSHSPSPGNPGTMIPCNCSAWPSPGLQPPLAYILLPGMGKPQLDPYPAAYAAAL
                   "

RESULT 3

BASE COUNT   1996 a   2143 c   2104 g   2124 t   5 others
ORIGIN

Query Match   Best Local Similarity   100.0%;   Score 500;   DB 9;   Length 8372;
Matches 500;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
Pred. No. 1.8e-144;

```
Qy      1 AAGCTTGGTGTGCCATCTATTTTGGACTATGCCTGCATACAGCTTATGGGAACATTTGTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTGGTGTGCCATCTATTTTGGACTATGCCTGCATACAGCTTATGGGAACATTTGTC 60

Qy     61 AGGCAAAAGTATAATAATGGCAAACTCTACGCCTTTTATTTTAAATTAGATTGGTGTGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGCAAAAGTATAATAATGGCAAACTCTACGCCTTTTATTTTAAATTAGATTGGTGTGAT 120

Qy    121 TTGATGCTGACGGGAGTGAGAGTAATGGCCTTATCCTGCTGCAGGCCTGTGCTGAGGATGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGATGCTGACGGGAGTGAGAGTAATGGCCTTATCCTGCTGCAGGCCTGTGCTGAGGATGG 180

Qy    181 CCTGGTCTCTGCCACCCTCCTCGAGTAGCCATTTTGCATGTGTAACAGGGGTCTCCCCTCTGGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTGGTCTCTGCCACCCTCCTCGAGTAGCCATTTTGCATGTGTAACAGGGGTCTCCCCTCTGGG 240

Qy    241 GCACAACAACAAAAGAGAAGAGTTGCTAAGGACAAGACAGGTGCGGAAATGCATCTCCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCACAACAACAAAAGAGAAGAGTTGCTAAGGACAAGACAGGTGCGGAAATGCATCTCCAT 300

Qy    301 TGGAACAGCCCTGGGCTTACTCCAATGGCTGAGAGGTGGCTATGGCCAGTCCTCCCCAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGAACAGCCCTGGGCTTACTCCAATGGCTGAGAGGTGGCTATGGCCAGTCCTCCCCAGA 360

Qy    361 GCTCTGCAGCTGCACTTTGGGGGGTGGAGGTGTCCTGCTTGTCGTCGTGTGTCGTGATAACGGCCGT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTCTGCAGCTGCACTTTGGGGGGTGGAGGTGTCCTGCTTGTCGTCGTGTGTCGTGATAACGGCCGT 420

Qy    421 GAAAGCCAGCCAACTGCTGCCCAAAATCACCCAGCCGATTGGGGGTTTCCCATCGGCGCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAAGCCAGCCAACTGCTGCCCAAAATCACCCAGCCGATTGGGGGTTTCCCATCGGCGCA 480

Qy    481 CCCTGCCCCGGGAGCCAAGAAG 500
          ||||||||||||||||||||||
Db    481 CCCTGCCCCGGGAGCCAAGAAG 500
```

RESULT 3

LOCUS   AL159970   112929 bp   DNA   linear   PRI 26-FEB-2001
AL159970/c
DEFINITION   Human DNA sequence from clone RP11-140119 on chromosome 13 Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL159970
VERSION   AL159970.16   GI:11121011
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112929)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:10715832.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-140I19 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-140I19 is at 112929 in this
            sequence. The true right end of clone RP11-477B16 is at 100 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            RP11-140I19 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..112929
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-140I19"
                     /clone_lib="RPCI-11.1"
     misc_feature    complement(1..86)
                     /note="match: GSS: Em:AQ036762"
     misc_feature    109..468
                     /note="match: GSS: Em:AQ894459"
     repeat_region   466..529
                     /note="match: GSS: Em:AQ894459"
     misc_feature    477..589
                     /note="16 copies 4 mer atct 95% conserved"
     repeat_region   564..1497
                     /note="match: GSS: Em:AQ894459"
     repeat_region   1813..2011
                     /note="L1ME repeat: matches 4782..5719 of consensus"
     repeat_region   2002..3064
                     /note="L1M4 repeat: matches 4138..4360 of consensus"
     repeat_region   3386..3984
                     /note="L1PA5 repeat: matches 5077..6143 of consensus"
     repeat_region   4276..4425
                     /note="L1M4 repeat: matches 3167..3764 of consensus"
     repeat_region   4443..4644
                     /note="L1MC5 repeat: matches 7695..7860 of consensus"
     repeat_region   4645..4937
                     /note="AluSx repeat: matches 1..202 of consensus"
     misc_feature    complement(4901..5044)
                     /note="L1MC5 repeat: matches 7013..7296 of consensus"
     repeat_region   5290..5333
                     /note="match: GSS: Em:AQ826245"
     repeat_region   5336..5567
                     /note="11 copies 4 mer tcta 79% conserved"

     repeat_region   6123..6165
                     /note="8 copies 4 mer atct 90% conserved"
     repeat_region   6297..6445
                     /note="L2 repeat: matches 2708..2749 of consensus"
     repeat_region   6863..7056
                     /note="MER5A repeat: matches 15..188 of consensus"
     repeat_region   7077..7301
                     /note="MER58B repeat: matches 115..325 of consensus"
     repeat_region   7302..7375
                     /note="AluY repeat: matches 1..296 of consensus"
     repeat_region   7763..7896
                     /note="MER58B repeat: matches 1..75 of consensus"
     repeat_region   8098..8267
                     /note="AluSq repeat: matches 119..299 of consensus"
     repeat_region   8288..8366
                     /note="AluJo/FRAM repeat: matches 164..302 of consensus"
     repeat_region   8387..8555
                     /note="L1ME repeat: matches 5707..5792 of consensus"
     misc_feature    8500..8868
                     /note="MIR repeat: matches 56..253 of consensus"
     repeat_region   9776..9958
                     /note="L2 repeat: matches 2563..2750 of consensus"
     misc_feature    10450..11039
                     /note="match: GSS: Em:AQ387731"
     misc_feature    12736..13078
                     /note="match: GSS: Em:AZ387316"
     misc_feature    complement(12818..13636)
                     /note="match: GSS: Em:AQ780981"
     misc_feature    13633..14078
                     /note="match: GSS: Em:AQ121010"
     repeat_region   14068..14346
                     /note="AluJb repeat: matches 19..296 of consensus"
     repeat_region   15874..16333
                     /note="ML1C repeat: matches 4..441 of consensus"
     repeat_region   17826..18033
                     /note="AluSg/x repeat: matches 102..303 of consensus"
     misc_feature    18733..19156
                     /note="match: GSS: Em:B40697"
     repeat_region   18773..18900
                     /note="AluJb repeat: matches 158..299 of consensus"
     repeat_region   19424..22318
                     /note="L2 repeat: matches 2695..2733 of consensus"
     repeat_region   22642..22812
                     /note="L1B1 repeat: matches 3058..6155 of consensus"
     misc_feature    complement(23696..24120)
                     /note="match: GSS: Em:AQ432849"
     repeat_region   26162..26358
                     /note="L2 repeat: matches 1806..1987 of consensus"
     repeat_region   26362..26483
                     /note="MIR repeat: matches 92..218 of consensus"
     repeat_region   26484..26792
                     /note="AluSx repeat: matches 1..307 of consensus"
     repeat_region   26793..26831
                     /note="AluJo repeat: matches 1..307 of consensus"
     repeat_region   29922..30027
                     /note="MIR repeat: matches 43..156 of consensus"
     repeat_region   30028..30311
                     /note="AluJb repeat: matches 1..303 of consensus"
     repeat_region   30312..30394
                     /note="MIR repeat: matches 156..236 of consensus"
     repeat_region   30500..30822
                     /note="AluSx repeat: matches 2..309 of consensus"
     repeat_region   30924..31008
                     /note="AluS repeat: matches 2..86 of consensus"
     repeat_region   31828..32083
                     /note="AluSq repeat: matches 1..257 of consensus"
```

```
repeat_region    33974..34098
                 /note="L2 repeat: matches 2632..2750 of consensus"
repeat_region    34233..34594
                 /note="THE1B repeat: matches 1..364 of consensus"
repeat_region    35032..35151
                 /note="MIR repeat: matches 13..139 of consensus"
repeat_region    35901..36022
                 /note="L1M3 repeat: matches 7611..7739 of consensus"
repeat_region    36470..37357
                 /note="L1M3 repeat: matches 5211..6137 of consensus"
repeat_region    37907..37952
                 /note="18 copies 4 mer agag 69% conserved"
repeat_region    44041..44094
                 /note="23 copies 2 mer ag 83% conserved"
misc_feature     complement(45963..46422)
                 /note="27 copies 2 mer ag tt 73% conserved"
repeat_region    41568..41873
                 /note="L1PA13 repeat: matches 5811..6156 of consensus"
misc_feature     complement(42666..43336)
                 /note="match: GSS: Em:AQ587521"
repeat_region    44021..44092
                 /note="MIR repeat: matches 50..181 of consensus"
repeat_region    50616..50671
                 /note="L1M84 repeat: matches 5417..5472 of consensus"
repeat_region    50691..50781
                 /note="AluJ/FLAM repeat: matches 2..94 of consensus"
repeat_region    50779..50885
                 /note="AluJb repeat: matches 1..111 of consensus"
repeat_region    49585..49869
                 /note="AluJo repeat: matches 1..284 of consensus"
repeat_region    50126..50429
                 /note="AluSg repeat: matches 1..305 of consensus"
repeat_region    50481..50601
                 /note="L1M84 repeat: matches 5485..5919 of consensus"
repeat_region    51254..51694
                 /note="AluJb repeat: matches 120..311 of consensus"
repeat_region    51064..51253
                 /note="AluSq repeat: matches 1..309 of consensus"
repeat_region    51718..52053
                 /note="L1M84 repeat: matches 30..365 of consensus"
repeat_region    52057..52298
                 /note="MIR1A1 repeat: matches 5928..6180 of consensus"
repeat_region    52441..52496
                 /note="L1M84 repeat: matches 5928..6180 of consensus"

misc_feature     /note="match: GSS: Em:AQ108740"
```

```
Query Match       89.2%;  Score 445.8; DB 9;  Length 112929;
Best Local Similarity  98.5%;   Pred. No. 2.1e-127;
Matches   450;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy       1 AAGCTTGGTGCCATCTATTTTGGACTATATGTCCTTGCTGCTTGGGGACAGTGTCGTGGTGCCTTGTGCCTTGGGGATAACGGCCGT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42912 AAGCTTGGTGCCATCTATTTTGGACTATATGTCCTTGCTGCTTGGGGACAGTGTCGTGGTGCCTTGTGCCTTGGGGATAACGGCCGT 42853

Qy      61 AGGCAAAAAGTATAATAAATGGCAAACTCTACGCCCTTTTATTTTTTAAATTAGATTGGTGTGGAT 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42852 AGGCAAAAAGTATAATAAATGGCAAACTCTACGCCCTTTTATTTTTTAAATTAGATTGGTGTGGAT 42793

Qy     121 TTGATGTGCTGACGGGAGGTAATGGCCTTATCCTGCTGCAGGCCTGCTGTGCTGAGGATGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42792 TTGATGTGCTGACGGGAGGTAATGGCCTTATCCTGCTGCAGGCCTGCTGTGCTGAGGATGG 42733

Qy     181 CCTGGTGTGCCACCCTCTCGAGTAGCATTTTGCATGTGTAACAGGTGTCTCCCCTCTGGG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42732 CCTGGTGTGCCACCCTCTCGAGTAGCATTTTGCATGTGTAACAGGTGTCTCCCCTCTGGG 42673

Qy     241 GCACAACAAACAAAGAGAACAGAGAGTTGCTAAGGACAAGAGAGCAGGTGCGGAAATGCATCTCCCAT 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42672 GCACAACAAACAAAGAGAACAGAGAGTTGCTAAGGACAAGAGAGCAGGTGCGGAAATGCATCTCCCAT 42613

Qy     301 TGGAACAGCCCTGGGCTTACTCCAATGGCTATGGCCAGTGCTCTCCCAGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42612 TGGAACAGCCCTGGGCTTACTCCAATGGCTATGGCCAGTGCTCTCCCAGA 42553
```

```
Qy     361 GCTCTCGCCAGTGTGCCACTTGGGGTGGGGGGGACAGTGTCGTGGTGCCTTGTGCCTTGTGCTGCTGATAACGGCCGT 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42552 GCTCTCGCCAGTGTGCCACTTGGGGTGGGGGGGACAGTGTCGTGGTGCCTTGTGCCTTGTGCTGCTGATAACGGCCGT 42493

Qy     421 GAAAGCCAGCCAAACTGCTGCCCAAAATCACCCAGCCG 457
           |||||||||||||||||||||||||||||||||||| |||
Db   42492 GAAAGCCAGCCAAACTGCTGCCCAGCAACGCAGAGCTG 42456
```

```
RESULT 4
AK055459/c
LOCUS       AK055459                2984 bp    mRNA            linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ30897 fis, clone FEBRA2005476.
ACCESSION   AK055459
VERSION     AK055459.1  GI:16550186
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens fetus brain cDNA to mRNA, clone_lib:FEBRA2
            clone:FEBRA2005476.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Nishi,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 2984)
AUTHORS     Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1..2984
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="FEBRA2005476"
                     /tissue_type="brain"
                     /clone_lib="FEBRA2"
                     /dev_stage="fetus"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      773 a    661 c    703 g    847 t
ORIGIN
```

```
Query Match       11.6%;  Score 58;  DB 9;  Length 2984;
Best Local Similarity 100.0%;   Pred. No. 9.6e-07;
Matches    58;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     443 AAAATCACCCAGCCGATTGGGGGGTTTGGGGGTTTCCCATCGGCGGCCACCCTGCCCCGGAGCCCAAGAAG 500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAATCACCCAGCCGATTGGGGGGTTTCCCATCGGCGGCCACCCTGCCCCGGAGCCCAAGAAG 4
```

```
RESULT 5
AL137061/c
LOCUS       AL137061              180538 bp    DNA             linear   PRI 01-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-477B16 on chromosome 13,
            complete sequence.
ACCESSION   AL137061
```

```
VERSION     AL137061.12  GI:11034478
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180538)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 26, 2000 this sequence version replaced gi:9799955.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-477B16 is from the library RPCI-11 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-477B16 The true
            left end of clone RP11-140119 is at 124009 in this sequence. The
            true right end of clone RP11-171014 is at 68026 in this sequence.
FEATURES             Location/Qualifiers
     source          1..180538
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-477B16"
                     /clone_lib="RPCI-11.2"
     repeat_region   1..992
     repeat_region   1029..2800
                     /note="L1MD2 repeat: matches 5086..6127 of consensus"
     repeat_region   2991..3288
                     /note="L1PA5 repeat: matches 4364..6141 of consensus"
     repeat_region   3398..3704
                     /note="AluSx repeat: matches 1..298 of consensus"
     repeat_region   4400..4423
                     /note="AluSx repeat: matches 2..310 of consensus"
     misc_feature    4827..5526
                     /note="12 copies 2 mer tt 100% conserved"
     repeat_region   5817..5867
                     /note="match: GSS: Em:AQ587522"
     repeat_region   6031..6270
                     /note="17 copies 3 mer gag 72% conserved"
     repeat_region   6271..6444
                     /note="L1PA5 repeat: matches 4974..5233 of consensus"
     repeat_region   6445..6522
                     /note="MER5B repeat: matches 1..178 of consensus"
     repeat_region   6816..7120
                     /note="L1PA5 repeat: matches 4901..4974 of consensus"
     repeat_region   7121..7160
                     /note="AluSq repeat: matches 1..303 of consensus"
     repeat_region   7304..7605
                     /note="20 copies 2 mer tt 77% conserved"
```

```
     repeat_region   7739..7950
                     /note="AluSq repeat: matches 1..302 of consensus"
     repeat_region   8363..8479
                     /note="L1ME3 repeat: matches 5937..6161 of consensus"
     repeat_region   8709..8797
                     /note="FLAM_C repeat: matches 1..118 of consensus"
     repeat_region   10879..11390
                     /note="AluJo repeat: matches 32..295 of consensus"
     repeat_region   11404..11694
                     /note="L2 repeat: matches 2209..2709 of consensus"
     repeat_region   11697..11931
                     /note="AluSx repeat: matches 1..291 of consensus"
     repeat_region   11940..12157
                     /note="AluJb repeat: matches 183..311 of consensus"
     misc_feature    11958..12368
                     /note="L1ME1E repeat: matches 180..419 of consensus"
     repeat_region   12237..12408
                     /note="match: GSS: Em:AQ605014"
     repeat_region   14529..14826
                     /note="L1T1E repeat: matches 1..68 of consensus"
     repeat_region   14848..14962
                     /note="AluSx repeat: matches 1..299 of consensus"
     repeat_region   15216..15303
                     /note="5S repeat: matches 9..119 of consensus"
     repeat_region   15427..15479
                     /note="L1ME3A repeat: matches 2 mer ta 72% conserved"
     repeat_region   16252..16494
                     /note="MIR repeat: matches 61..116 of consensus"
     repeat_region   16557..17174
                     /note="L1MC/D repeat: matches 5279..5553 of consensus"
     repeat_region   17179..17204
                     /note="L1MC5 repeat: matches 6906..7534 of consensus"
     repeat_region   17250..17554
                     /note="13 copies 2 mer aa 92% conserved"
     repeat_region   17587..17639
                     /note="AluSx repeat: matches 3..308 of consensus"
     repeat_region   17640..17940
                     /note="L1MC5 repeat: matches 7603..7657 of consensus"
     repeat_region   17941..18210
                     /note="AluSx repeat: matches 1..301 of consensus"
     repeat_region   18499..18594
                     /note="L1MC5 repeat: matches 7657..7909 of consensus"
     repeat_region   18602..18952
                     /note="L1ME3A repeat: matches 5786..5888 of consensus"
     repeat_region   18953..19259
                     /note="Tigger4(Zombi) repeat: matches 1..345 of consensus"
     repeat_region   19260..19813
                     /note="AluSp repeat: matches 1..308 of consensus"
     repeat_region   19818..20095
                     /note="Tigger4(Zombi) repeat: matches 345..919 of
                     consensus"
     repeat_region   20096..20141
                     /note="AluJo repeat: matches 1..273 of consensus"
     repeat_region   20145..20886
                     /note="23 copies 2 mer aa 76% conserved"
     repeat_region   20918..21065
                     /note="Tigger4(Zombi) repeat: matches 905..1647 of
                     consensus"
     repeat_region   21879..22171
                     /note="L1ME3A repeat: matches 5514..5661 of consensus"
     repeat_region   22565..22586
                     /note="AluSx repeat: matches 1..288 of consensus"
     repeat_region   22673..22825
                     /note="11 copies 2 mer ag 100% conserved"
     repeat_region   22716..23202
                     /note="MER69A repeat: matches 4..175 of consensus"
     misc_feature    22716..23202
                     /note="match: GSS: Em:AQ695208"
```

```
RESULT 6

Db    76225  AC005369
Qy      443

LOCUS       AC005369   74371 bp   DNA   linear   PRI 01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 11j3 (LBNL H175), complete
            sequence.
ACCESSION   AC005369
VERSION     AC005369.1  GI:3367505
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 74371)
```

```
AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
          Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
          Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE   Sequencing of human chromosome 5
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 74371)
  AUTHORS Ricke,D.O.
  TITLE   Large Scale Sequence Analysis and Annotation with the Sequence
          Comparison Analysis (SCAN) System
  JOURNAL Unpublished
REFERENCE 3  (bases 1 to 74371)
  AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
          Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
          Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE   Direct Submission
  JOURNAL Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
          Institute, Lawrence Berkeley National Laboratory, MS 74-157,
          Berkeley, CA 94720, U.S.A.
COMMENT   Sequence submitted by:
          DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1..74371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q"
                     /clone="11j3"
                     /note="LBNL H175"
     repeat_region   893..1030
                     /rpt_family="Alu"
     repeat_region   2295..2438
                     /rpt_family="Alu"
     repeat_region   2818..2859
                     /note="(GT)21"
                     /rpt_type=tandem
                     /rpt_unit=GT
     misc_feature    join(3246..3410,3721..3828)
                     /note="658 & 698 protein identity GenPept:U22377"
                     /standard_name="RLF"
     repeat_region   3431..3724
                     /rpt_family="Alu"
     repeat_region   3707..3728
                     /note="(A)22"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   4366..4661
                     /rpt_family="Alu"
     repeat_region   5327..5602
                     /rpt_family="Alu"
     repeat_region   6586..6956
                     /rpt_family="L1"
     repeat_region   6647..6684
                     /note="(CA)19"
                     /rpt_type=tandem
                     /rpt_unit=CA
     repeat_region   7113..7373
                     /rpt_family="Alu"
     repeat_region   8258..8503
                     complement(7830..8185)
                     /standard_name="possible repeat"
     repeat_region   9070..9387
                     /rpt_family="Alu"
     repeat_region   complement(9740..9845)
                     /rpt_family="Alu"
     repeat_region   complement(10440..11015)
                     /rpt_family="MER42"
     repeat_region   11950..12250
                     /rpt_family="Alu"
     repeat_region   12057..12085
                     /note="(A)29"
                     /rpt_type=tandem
                     /rpt_unit=A
```

```
     misc_feature    complement(25060..25580)
                     /note="match: GSS: Em:AZ378801"
     repeat_region   26168..26343
                     /note="88 copies 2 mer tt 58% conserved"
     repeat_region   30377..30404
                     /note="14 copies 2 mer ca 92% conserved"
     repeat_region   31111..31420
                     /note="AluJo repeat: matches 5..312 of consensus"
     repeat_region   31844..31858
                     /note="MLT1C repeat: matches 3..17 of consensus"
     repeat_region   31859..32197
                     /note="THE1B repeat: matches 1..364 of consensus"
     repeat_region   32198..32655
                     /note="MLT1C repeat: matches 17..466 of consensus"
     repeat_region   33159..33234
                     /note="MLT1A1 repeat: matches 1..288 of consensus"
     repeat_region   33235..33531
                     /note="MLT1A1 repeat: matches 288..365 of consensus"
     repeat_region   33532..33830
                     /note="AluSx repeat: matches 3..298 of consensus"
     repeat_region   34287..34322
                     /note="MLT1A1 repeat: matches 1..288 of consensus"
     repeat_region   34337..34440
                     /note="MER53 repeat: matches 154..189 of consensus"
     repeat_region   34374..34417
                     /note="7 copies 4 mer gtgt 96% conserved"
     repeat_region   34441..34739
                     /note="26 copies 2 mer ct 75% conserved"
     repeat_region   34830..34955
                     /note="AluSx repeat: matches 1..311 of consensus"
     repeat_region   35551..35828
                     /note="L2 repeat: matches 2421..2549 of consensus"
     repeat_region   36576..36879
                     /note="L2 repeat: matches 1687..2326 of consensus"
     repeat_region   37816..37843
                     /note="AluY repeat: matches 1..303 of consensus"
     misc_feature    39198..39249
                     /note="match: GSS: Em:AQ781559"
     repeat_region   39326..40065
                     /note="MIR repeat: matches 196..251 of consensus"
     repeat_region   39933..39974
                     /note="21 copies 2 mer aa 78% conserved"
     repeat_region   41127..41702
                     /note="L2 repeat: matches 2326 of consensus"
     repeat_region   43562..43601
                     /note="20 copies 2 mer ac 100% conserved"
     repeat_region   43778..44085
                     /note="AluSx repeat: matches 4..310 of consensus"
     repeat_region   44763..44790
                     /note="7 copies 4 mer gtgt 96% conserved"
```

```
Query Match   11.6%;   Score 58;   DB 9;   Length 180538;
Best Local Similarity  100.0%;   Pred. No. 1.8e-06;
Matches  58;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy      443 AAAATCACCCAGCCGAGTTGGGGGGGTTTCCCATCGGCGGCACCCTGCCCCGGAGCCAAGAAG 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    76225 AAAATCACCCAGCCGAGTTGGGGGGGTTTCCCATCGGCGGCACCCTGCCCGGAGCCAAGAAG 76168
```

repeat_region   12365..12645
    /rpt_family="Alu"
repeat_region   13727..13750
    /note="(AC)12"
    /rpt_unit=AC
    /rpt_type=tandem
repeat_region   13783..14024
    /rpt_family="L1"
repeat_region   14175..14470
    /rpt_family="Alu"
misc_feature   complement(14906..15259)
    /rpt_family="Alu"
    /standard_name="possible repeat"
repeat_region   15300..15613
    /rpt_family="Alu"
repeat_region   16671..16690
    /note="(T)20"
    /rpt_unit=T
    /rpt_type=tandem
repeat_region   complement(16675..16977)
    /rpt_family="Alu"
repeat_region   complement(16993..17085)
    /rpt_family="MER42"
repeat_region   complement(17678..18276)
    /rpt_family="Alu"
repeat_region   19305..19583
    /rpt_family="Alu"
repeat_region   19914..19945
    /note="(GTTT)8"
    /rpt_unit=GTTT
    /rpt_type=tandem
repeat_region   21202..21496
    /rpt_family="Alu"
repeat_region   21736..22035
    /rpt_family="Alu"
repeat_region   22017..22038
    /note="(A)22"
    /rpt_unit=A
    /rpt_type=tandem
repeat_region   22221..22457
    /rpt_family="Alu"
repeat_region   complement(22638..22981)
    /standard_name="possible repeat"
repeat_region   23022..23326
    /rpt_family="Alu"
repeat_region   23473..23761
    /rpt_family="Alu"
repeat_region   23744..23767
    /note="(A)24"
    /rpt_unit=A
    /rpt_type=tandem
repeat_region   complement(24639..24694)
    /rpt_family="MER42"
repeat_region   complement(25349..25713)
    /rpt_family="Alu"
repeat_region   complement(25727..26471)
    /rpt_family="Alu"
repeat_region   27191..27477
    /rpt_family="Alu"
repeat_region   27774..28057
    /rpt_family="Alu"
repeat_region   28040..28066
    /note="(A)27"
    /rpt_unit=A
    /rpt_type=tandem
repeat_region   28712..28930
    /rpt_family="MER20"
misc_feature   complement(28769..28838)
    /note="GRAIL 2 excellent exon, frame 0"
repeat_region   28987..29214
    /rpt_family="Alu"
misc_feature   join(29485..29559,29598..29779)

---

misc_feature   29495..29976
    /note="95% identity dbSTS:G14522 (SHGC-11312)"
    /standard_name="AI027942"
misc_feature   complement(30401..30536)
    /note="100% identity EST ov84a10.x1"
repeat_region   complement(30682..30733)
    /rpt_family="MIR"
misc_feature   complement(31573..31724)
    /note="GRAIL 2 excellent exon, frame 1"
misc_feature   complement(32159..32232)
    /note="GRAIL 2 excellent exon, frame 2"
repeat_region   32388..32488
    /rpt_family="MLT1"
repeat_region   32617..32908
    /rpt_family="Alu"
repeat_region   32977..33088
    /rpt_family="Alu"
repeat_region   complement(33670..33785)
    /rpt_family="MLT1"
misc_feature   complement(34021..34144)
    /note="GRAIL 2 excellent exon, frame 2"
misc_feature   complement(35238..35331)
    /note="GRAIL 2 excellent exon, frame 1"
repeat_region   36392..36663
    /rpt_family="Alu"
misc_feature   36901..37222
    /note="100% identity EST ou55c09.x1"
    /db_xref="dbEST:AI025011"
STS   36901..37164
    /standard_name="STSG-9983"
    /db_xref="dbSTS:G26554"
misc_feature   complement(37595..37654)
    /note="GRAIL 2 excellent exon, frame 0"
misc_feature   complement(join(38063..38218,38462..38578,38741..38995,
                 39071..39205,39532..39630,39935..40048,40300..40410,
                 40503..40661,41868..41972,42103..42225,42492..42569,
                 44379..44507))
    /standard_name="histidyl-tRNA synthetase"
    /note="78%-100% protein identity GenPept:U18937"
misc_feature   complement(38069..38215)
    /note="GRAIL 2 excellent exon, frame 0"
misc_feature   complement(38462..38578)

Qy   46 ATGGGAACATTTGTCAGGCAAAAGTATAATAATGGCAAAACTCTACGGCCCTTTTATTTAAAA 105
           |: ::::: ::| :: ||| :|| ::::::| ||::::: |:: ||||| ||||
Db   15844 AWTKSMRRMMWWWTWTWWWWWWNRRAAAAWTTTWARKKRAAAWYCAYAYAYWYTTTTTTTTTTTTTTT 15903

Qy   106 TTAGATTGGTGTGTGATTTGCATGCGTGCACGGGAGTGGAGAGTTGGAGAGTAATGGCCTTATCCTGCTGCAGG 165
           |::||| :: ::: :| |:|||:|:| |:|:::|:| ::: :|:
Db   15904 TKWGASRSRGWSTCKCKCTSTSKCSCMSRSKSKRGWGYRSWKKYRCAMMWTCKSSKCWCW 15963

Qy   166 CTGTGCTGTGAGGATGGCCTGGTCTGCCACCCTCCTGCGAGTAGCATTTTGCATGTGTAACAG 225
           |: : :::::|::| |::| ::: :| :::::| :: :|| ||||::|:|
Db   15964 SYRMRMKCYSCSYCYCSSGKKYNCRCSMYWYTCYSYKXYWSMSYCTCTSWGWRWMWNSKG 16023

Qy   226 GGTCTCCCCTCTGGGGCACAACAACAAACAAGAGAAGTT 261
           ||:: :::: |::|| | |:| :: |: ||
Db   16024 RSWMYASRSGCSCGSCSMCMCRCSCSMSMKMWWWTTT 16059

Query Match   8.8%;   Score 44;   DB 9;   Length 74371;
Best Local Similarity   16.2%;   Pred. No. 0.036;
Matches   35;   Conservative 98; Mismatches 83; Indels 0; Gaps 0;

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 147124)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens, clone RP11-338H7
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 147124)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 1, 2000 this sequence version replaced gi:6922528.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu

            Project Information
            Center project name: L5860
            Center clone name: 338_H_7

            NOTE: This record contains 156 individual
*   sequencing reads that have not been assembled into
*   contigs. Runs of N are used to separate the reads
*   and the order in which they appear is completely
*   arbitrary. Low-pass sequence sampling is useful for
*   identifying clones that may be gene-rich and allows
*   overlap relationships among clones to be deduced.
*   However, it should not be assumed that this clone
*   will be sequenced to completion. In the event that
*   the record is updated, the accession number will
*   be preserved.
*
*   1   793: contig of 793 bp in length
*   794   893: gap of   100 bp
*   894   1668: contig of 775 bp in length
*   1669   1768: gap of   100 bp
*   1769   2609: contig of 841 bp in length
*   2610   2709: gap of   100 bp
*   2710   3509: contig of 800 bp in length
*   3510   3609: gap of   100 bp
*   3610   4428: contig of 819 bp in length
*   4429   4528: gap of   100 bp
*   4529   5307: contig of 779 bp in length
*   5308   5407: gap of   100 bp
*   5408   6197: contig of 790 bp in length
*   6198   6297: gap of   100 bp
*   6298   6986: contig of 689 bp in length
*   6987   7086: gap of   100 bp
*   7087   7838: contig of 752 bp in length
*   7839   7938: gap of   100 bp
*   7939   8707: contig of 769 bp in length
*   8708   8807: gap of   100 bp
*   8808   9603: contig of 796 bp in length
*   9604   9703: gap of   100 bp
*   9704 10510: contig of 807 bp in length
*   10511 10610: gap of   100 bp
*   10611 11388: contig of 778 bp in length
*   11389 11488: gap of   100 bp
*   11489 12295: contig of 807 bp in length
*   12296 12395: gap of   100 bp
*   12396 13230: contig of 835 bp in length
*   13231 13330: gap of   100 bp
*   13331 14176: contig of 846 bp in length
*   14177 14276: gap of   100 bp
*   14277 15045: contig of 769 bp in length
*   15046 15145: gap of   100 bp
*   15146 15917: contig of 772 bp in length
*   15918 16017: gap of   100 bp
*   16018 16777: contig of 760 bp in length
*   16778 16877: gap of   100 bp
*   16878 17673: contig of 796 bp in length
*   17674 17773: gap of   100 bp
*   17774 18524: contig of 751 bp in length
*   18525 18624: gap of   100 bp
*   18625 19423: contig of 799 bp in length
*   19424 19523: gap of   100 bp
*   19524 20326: contig of 803 bp in length
*   20327 20426: gap of   100 bp
*   20427 21202: contig of 776 bp in length
*   21203 21302: gap of   100 bp
*   21303 22069: contig of 767 bp in length
*   22070 22169: gap of   100 bp
*   22170 22971: contig of 802 bp in length
*   22972 23071: gap of   100 bp
*   23072 23876: contig of 805 bp in length
*   23877 23976: gap of   100 bp
*   23977 24774: contig of 798 bp in length
*   24775 24874: gap of   100 bp
*   24875 25650: contig of 776 bp in length
*   25651 25750: gap of   100 bp
*   25751 26542: contig of 792 bp in length
*   26543 26642: gap of   100 bp
*   26643 27433: contig of 791 bp in length
*   27434 27533: gap of   100 bp
*   27534 28325: contig of 792 bp in length
*   28326 28425: gap of   100 bp
*   28426 29235: contig of 810 bp in length
*   29236 29335: gap of   100 bp
*   29336 30134: contig of 799 bp in length
*   30135 30234: gap of   100 bp
*   30235 31044: contig of 810 bp in length
*   31045 31144: gap of   100 bp
*   31145 31923: contig of 779 bp in length
*   31924 32023: gap of   100 bp
*   32024 32804: contig of 781 bp in length
*   32805 32904: gap of   100 bp
*   32905 33648: contig of 744 bp in length
*   33649 33748: gap of   100 bp
*   33749 34550: contig of 802 bp in length
*   34551 34650: gap of   100 bp
*   34651 35398: contig of 748 bp in length
*   35399 35498: gap of   100 bp
*   35499 36304: contig of 806 bp in length
*   36305 36404: gap of   100 bp
*   36405 37202: contig of 798 bp in length
*   37203 37302: gap of   100 bp
*   37303 38139: contig of 837 bp in length
*   38140 38239: gap of   100 bp
*   38240 39045: contig of 806 bp in length
*   39046 39145: gap of   100 bp
*   39146 39931: contig of 786 bp in length
*   39932 40031: gap of   100 bp
*   40032 40830: contig of 799 bp in length
*   40831 40930: gap of   100 bp
*   40931 41701: contig of 771 bp in length
*   41702 41801: gap of   100 bp

```
*  41802  42585: contig of 784 bp in length
*  42586  42685: gap of 100 bp
*  42686  43490: contig of 805 bp in length
*  43491  43590: gap of 100 bp
*  43591  44360: contig of 770 bp in length
*  44361  44460: gap of 100 bp
*  44461  45263: contig of 803 bp in length
*  45264  45363: gap of 100 bp
*  45364  46146: contig of 783 bp in length
*  46147  46246: gap of 100 bp
*  46247  47065: contig of 819 bp in length
*  47066  47165: gap of 100 bp
*  47166  47956: contig of 791 bp in length
*  47957  48056: gap of 100 bp
*  48057  48833: contig of 777 bp in length
*  48834  48933: gap of 100 bp
*  48934  49716: contig of 783 bp in length
*  49717  49816: gap of 100 bp
*  49817  50577: contig of 761 bp in length
*  50578  50677: gap of 100 bp
*  50678  51461: contig of 784 bp in length
*  51462  51561: gap of 100 bp
*  51562  52251: contig of 790 bp in length
*  52252  52351: gap of 100 bp
*  52352  53218: contig of 767 bp in length
*  53219  53318: gap of 100 bp
*  53319  54114: contig of 796 bp in length
*  54115  54214: gap of 100 bp
*  54215  55025: contig of 811 bp in length
*  55026  55125: gap of 100 bp
*  55126  55943: contig of 818 bp in length
*  55944  56043: gap of 100 bp
*  56044  56880: contig of 837 bp in length
*  56881  56980: gap of 100 bp
*  56981  57775: contig of 795 bp in length
*  57776  57875: gap of 100 bp
*  57876  58675: contig of 800 bp in length
*  58676  58775: gap of 100 bp
*  58776  59557: contig of 782 bp in length
*  59558  59657: gap of 100 bp
*  59658  60437: contig of 780 bp in length
*  60438  60537: gap of 100 bp
*  60538  61326: contig of 789 bp in length
*  61327  61426: gap of 100 bp
*  61427  62225: contig of 799 bp in length
*  62226  62325: gap of 100 bp
*  62326  63098: contig of 773 bp in length
*  63099  63198: gap of 100 bp
```

```
Query Match        55.1%;  Score 38.8;  DB 2;  Length 147124;
Best Local Similarity  55.1%;  Pred. No. 1.7;
Matches   76;  Conservative  0;  Mismatches  62;  Indels  0;  Gaps  0;

Qy     1 AAGCTTCGTGCTCATCTATTTTGGACTATGCCTTGCATACAGCTTTATGGGAACANATTGTC 60
Db 42442 AAGGCTTGGGCGCCACAGTATATATGTACAATGACTTGGATACATCTCAGAGGCATCANGTTGG 42383

Qy    61 AGGCAAAAAGTATATAATAGGGAAAACTCTACGCCTTTTTTTTTAAATTAGATTGGTGTGAT 120
Db 42382 ATGAAAAGAAGCTAATCTCAAAAAGAATATATATTCCATTTTTATTTGACTTTTTTGTTTTTGT 42323

Qy   121 TTGATGCTGACGGAGTG 138
Db 42322 TTTGTTTGTGAGGGTCTG 42305
```

RESULT 8
LOCUS   AK025967   1478 bp   mRNA   linear   PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22314 fis, clone HRC05250.
ACCESSION   AK025967
VERSION   AK025967.1 GI:10438649
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens primary human renal epithelial cells cDNA to mRNA, clone_lib:HRC clone:HRC05250.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1478)
AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE   Direct Submission
JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES   Location/Qualifiers
   source   1..1478
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_type="primary human renal epithelial cells"
            /clone="HRC05250"
            /clone_lib="HRC"
            /note="cloning vector pME18SFL3"
BASE COUNT 501 a 271 c 343 g 363 t
ORIGIN

```
Query Match        7.5%;  Score 37.4;  DB 9;  Length 1478;
Best Local Similarity  54.8%;  Pred. No. 2.3;
Matches   74;  Conservative  0;  Mismatches  61;  Indels  0;  Gaps  0;

Qy   146 TGGCCTTATCCTGCTGCACGGCTGTGCTGAGGATGGCCTGGTCTGCGACACCTTCCTCCAGTA 205
Db   422 TGGACTTATCCTCGGTGCCACTGAGACTGTTGCAGATTTAGATCTACAATCCACCTGGAATT 481

Qy   206 GCATTTTGCCATGTGTAACAGAGGGTGTCCCCTCGGGCACAACACAAAAGAGAAGTTTGCTA 265
Db   482 GATTTTTGTATATGGAGTGAGGTCCCACACTTACAGGGAAAAGCANCACCCGAAAGTGAGA 541

Qy   266 AGGACAAGAAGCAGG 280
Db   542 ATGCCTAGAGGCAGG 556
```

RESULT 9
LOCUS   AC023348/c   193046 bp   DNA   linear   HTG 21-SEP-2000
DEFINITION Homo sapiens chromosome RPCI-11 clone RP11-543B23, WORKING DRAFT SEQUENCE, 28 unordered pieces.
ACCESSION   AC023348
VERSION   AC023348.5 GI:10424452
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 193046)
AUTHORS   Waterston,R.H.
TITLE   The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 193046)

AUTHORS   Waterston, R.H.
TITLE   Direct Submission
JOURNAL   Submitted (11-FEB-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
COMMENT   On Sep 21, 2000 this sequence version replaced gi:9937829.

          -------------- Genome Center --------------
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc/index.shtml
          -------------- Project Information --------------
          Center project name: H_NH0543B23
          -------------- Summary Statistics --------------
          Sequencing vector: M13; 67%
          Sequencing vector: plasmid; 33%
          Chemistry: Dye-primer ET; 67% of reads
          Chemistry: Dye-terminator Big Dye; 33% of reads
          Assembly program: Phrap; version 0.990319
          Consensus quality: 169181 bases at least Q40
          Consensus quality: 177376 bases at least Q30
          Consensus quality: 182283 bases at least Q20
          Insert size: 190000; agarose-fp
          Insert size: 190346; sum-of-contigs
          Quality coverage: 3.33 in Q20 bases; agarose-fp
          Quality coverage: 3.39 in Q20 bases; sum-of-contigs

          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 28 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.

          *   1   1241: contig of   1241 bp in length
          *   1242   1341: gap of unknown length
          *   1342   2960: contig of   1619 bp in length
          *   2961   3060: gap of unknown length
          *   3061   4647: contig of   1587 bp in length
          *   4648   4747: gap of unknown length
          *   4748   6990: contig of   2243 bp in length
          *   6991   7090: gap of unknown length
          *   7091   11122: contig of   4032 bp in length
          *   11123   11222: gap of unknown length
          *   11223   14168: contig of   2946 bp in length
          *   14169   14268: gap of unknown length
          *   14269   17666: contig of   3398 bp in length
          *   17667   17766: gap of unknown length
          *   17767   21663: contig of   3897 bp in length
          *   21664   21763: gap of unknown length
          *   21764   27073: contig of   5310 bp in length
          *   27074   27173: gap of unknown length
          *   27174   31694: contig of   4521 bp in length
          *   31695   31794: gap of unknown length
          *   31795   36382: contig of   4588 bp in length
          *   36383   36482: gap of unknown length
          *   36483   39477: contig of   2995 bp in length
          *   39478   39577: gap of unknown length
          *   39578   45021: contig of   5444 bp in length
          *   45022   45121: gap of unknown length
          *   45122   49190: contig of   4069 bp in length
          *   49191   49290: gap of unknown length
          *   49291   53227: contig of   3937 bp in length
          *   53228   53327: gap of unknown length
          *   53328   57469: contig of   4142 bp in length
          *   57470   57569: gap of unknown length
          *   57570   63756: contig of   6187 bp in length
          *   63757   63856: gap of unknown length
          *   63857   72227: contig of   8271 bp in length
          *   72228   72327: gap of unknown length
          *   72328   77996: contig of   5769 bp in length
          *   77997   78096: gap of unknown length

          *   78097   83773: contig of   5677 bp in length
          *   83774   83873: gap of unknown length
          *   83874   92791: contig of   8918 bp in length
          *   92792   92891: gap of unknown length
          *   92892   104229: contig of 11338 bp in length
          *   104230   104329: gap of unknown length
          *   104330   116282: contig of 11953 bp in length
          *   116283   116382: gap of unknown length
          *   116383   128263: contig of 11881 bp in length
          *   128264   128363: gap of unknown length
          *   128364   141920: contig of 13557 bp in length
          *   141921   142020: gap of unknown length
          *   142021   156957: contig of 14937 bp in length
          *   156958   157057: gap of unknown length
          *   157058   173103: contig of 16046 bp in length
          *   173104   173203: gap of unknown length
          *   173204   193046: contig of 19843 bp in length.

FEATURES   Location/Qualifiers
     source   1..193046
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="RPCI-11"
                     /clone="RP11-543B23"
     misc_feature   1..1241
                     /note="assembly_name:Contig10"
     misc_feature   1342..2960
                     /note="assembly_name:Contig11"
     misc_feature   3061..4647
                     /note="assembly_name:Contig12"
     misc_feature   4748..6990
                     /note="assembly_name:Contig13"
     misc_feature   7091..11122
                     /note="assembly_name:Contig14"
     misc_feature   11223..14168
                     /note="assembly_name:Contig15"
     misc_feature   14269..17666
                     /note="assembly_name:Contig16"
     misc_feature   17767..21663
                     /note="assembly_name:Contig17"
     misc_feature   21764..27073
                     /note="assembly_name:Contig18"
     misc_feature   27174..31694
                     /note="assembly_name:Contig19"
     misc_feature   31795..35382
                     /note="assembly_name:Contig20"
     misc_feature   36483..39477
                     /note="assembly_name:Contig22"
     misc_feature   3958..45021
                     /note="assembly_name:Contig23"
     misc_feature   45122..49190
                     /note="assembly_name:Contig24"
     misc_feature   49291..53227
                     /note="assembly_name:Contig25"
     misc_feature   53328..57469
                     /note="assembly_name:Contig26"
     misc_feature   57570..63756
                     /note="assembly_name:Contig27"
     misc_feature   63857..72127
                     /note="assembly_name:Contig28"
     misc_feature   72228..77996
                     /note="assembly_name:Contig29"
     misc_feature   78097..83773
                     /note="assembly_name:Contig30"
     misc_feature   83874..92791
                     /note="assembly_name:Contig31"
     misc_feature   92892..104229
                     /note="assembly_name:Contig32"
     misc_feature   104330..116282
                     /note="assembly_name:Contig33"
     misc_feature   116383..128263
                     /note="assembly_name:Contig35"
     misc_feature   128364..141920

```
misc_feature        142021..156957
                    /note="assembly_name:Contig36"
misc_feature        157058..173103
                    /note="assembly_name:Contig37"
misc_feature        173204..193046
                    /note="assembly_name:Contig38"
BASE COUNT          56944 a   39057 c   37629 g   56658 t   2758 others
ORIGIN

Qy          39  CAGCTTTATGGGAACATTTGTCAGGCAAAAGTATAATAATAATGGCAAACTCTACGCCTTTTA  98
                |||||||||  ||| ||| || |||| | |  |||||||||||   ||||| ||| | ||||
Db       17291  CTGCTTTATTGCAAAGTTAGTTAGTTAGATCANAAATTAGAAATCATCACCCACTACTCTAAAGG  17232

Qy          99  TTTTAAATTAGATTGGTGTGATTTGATGCTGACGGGGAGTGAGAGTAATGGC  149
                ||-|||  ||| | |||||||| |||  |||| ||||| || |||| ||||||
Db       17231  GTTAAAATGTGCCTATGAGTGAATAGACTGAGAGGAGGAGTTACACGTGTTGGC  17181
```

RESULT 10
AC016716/c

LOCUS   AC016716   214269 bp   DNA   linear   PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-31213 from 2, complete sequence.
ACCESSION   AC016716
VERSION   AC016716.6 GI:11038573
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 214269)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   9063792
REFERENCE   2 (bases 1 to 214269)
  AUTHORS   Haakenson,B., Coblitz,B., Hawkins,M., Elliott,G., Abbott,A. and
            Boyer,E.
  TITLE   The sequence of Homo sapiens BAC clone RP11-31213
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 214269)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-OCT-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 214269)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 214269)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Oct 30, 2000 this sequence version replaced gi:9910084.
COMMENT   -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0312103

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping this
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An Improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR: pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-332A19; the clone sequenced
            to the right is RP11-11K6. Actual start of this clone is at base
            position 1 of RP11-11K6. Actual end is at base position 214269 of
            RP11-31213.

            There are polymorphic base pair differences in the overlap between
            the clone RP11-31213 and RP11-11K6.
```

FEATURES   Location/Qualifiers
     source   1..214269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-31213"
                     /clone_lib="RPCI-11"
     repeat_region   1078..1232
                     /rpt_family="MIR"
     repeat_region   1879..2058
                     /rpt_family="MIR"
     repeat_region   2125..2395
                     /rpt_family="MIR"
     repeat_region   2673..2750
                     /rpt_family="Alu"
     repeat_region   2783..2937
                     /rpt_family="MER1_type"
     misc_feature   3789..3864
                     /note="similar to   EST AW894776 (NID:g80588890)"
     repeat_region   3908..3960
                     /rpt_family="MIR"
     repeat_region   4536..4648
                     /rpt_family="Alu"
     repeat_region   6038..6206
                     /rpt_family="MIR"
     repeat_region   7287..7503
                     /rpt_family="MIR"
     repeat_region   7738..8004
                     /rpt_family="Alu"
     repeat_region   8081..8152
                     /rpt_family="MIR"
     repeat_region   8991..9042
                     /rpt_family="Mariner"
     repeat_region   9438..9623
                     /rpt_family="MIR"
     repeat_region   9894..10114
                     /rpt_family="L2"
     repeat_region   10292..10583
                     /rpt_family="MER1_type"

FEATURES
　　　source

```
Location/Qualifiers
1. .172177
```

RESULT 11
LOCUS　　　　　AL359377　　　　　　　　172177 bp　　　　DNA　　　　　linear　　PRI 15-NOV-2001
DEFINITION　 Human DNA sequence from clone RP11-96C4 on chromosome 10, complete
　　　　　　　 sequence.
ACCESSION　　 AL359377
VERSION　　　　AL359377.18　　GI:16972922
KEYWORDS　　　HTG.
SOURCE　　　　 human.
　ORGANISM　　 Homo sapiens
　　　　　　　 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　　 1　(bases 1 to 172177)
　AUTHORS　　　Brown,A.
　TITLE　　　　 Direct Submission
　JOURNAL　　　Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
　　　　　　　 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
　　　　　　　 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT　　　　On Nov 16, 2001 this sequence version replaced gi:16605669.
　　　　　　　 During sequence assembly data is compared from overlapping clones.
　　　　　　　 Where differences are found these are annotated as variations
　　　　　　　 together with a note of the overlapping clone name. Note that the
　　　　　　　 variation annotation may not be found in the sequence submission
　　　　　　　 corresponding to the overlapping clone, as we submit sequences with
　　　　　　　 only a small overlap as described above.
　　　　　　　 This sequence was finished as follows unless otherwise noted: all
　　　　　　　 regions were either double-stranded or sequenced with an alternate
　　　　　　　 chemistry or covered by high quality data (i.e., phred quality >=
　　　　　　　 30); an attempt was made to resolve all sequencing problems, such
　　　　　　　 as compressions and repeats; all regions were covered by at least
　　　　　　　 one plasmid subclone or more than one M13 subclone; and the
　　　　　　　 assembly was confirmed by restriction digest. The following
　　　　　　　 abbreviations are used to associate primary accession numbers given
　　　　　　　 in the feature table with their source databases: Em:, EMBL; Sw:,
　　　　　　　 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
　　　　　　　 database can be found at
　　　　　　　 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
　　　　　　　 was generated from part of bacterial clone contigs of human
　　　　　　　 chromosome 10, constructed by the Sanger Centre Chromosome 10
　　　　　　　 Mapping Group.　Further information can be found at
　　　　　　　 http://www.sanger.ac.uk/HGP/Chr10
　　　　　　　 RP11-96C4 is from the library RPCI-11.1 constructed by the group of
　　　　　　　 Pieter de Jong. For further details see
　　　　　　　 http://www.chori.org/bacpac/home.htm
　　　　　　　 VECTOR: pBACe3.6
　　　　　　　 IMPORTANT: This sequence is not the entire insert of clone
　　　　　　　 RP11-96C4 It may be shorter because we sequence overlapping
　　　　　　　 sections only once, except for a short overlap.
　　　　　　　 The true left end of clone RP11-384P3 is at 1 in this sequence.
　　　　　　　 The true left end of clone RP11-96C4 is at 170178 in this sequence.
　　　　　　　 The true right end of clone RP11-388P9 is at 29608 in this
　　　　　　　 sequence.

Qy　　39　CAGCTTTATGGGAACATTTGTCAGGCAAAAAGTATATAAATAGGCAAAACTCTACCCCTTTTA 98
　　　　　||||||||||　|　||||　|||　|||||||||　||||||||||||　|　|||||　|||||
Db　67598　CTGCTTTATTGCAAAGTTAGTTAGTTAGATCAAAATTAGAAATCATCCACCACTACTCTAAAGG 67539

Qy　　99　TTTTAAATTAGATTTGGTGTCTGTGATTTGATGTCTGACGGGGACTGAGAGTAAATGGC 149
　　　　　||||||||||||||　|||||||||　||||||||||||||||||　|||　||||　||||
Db　67538　GTTTAAARGTGCCTATGAGTGAATAGACTTGAGAAGGGAGTTACAGTGTTGGC 67488

Query Match　　　65;　Score 37.4;　DB 9;　Length 214269;
Best Local Similarity　58.6%;　Pred. No. 4.9;
Matches　65;　Conservative　0;　Mismatches　46;　Indels　0;　Gaps　0;

Qy　　39　CAGCTTTATGGGAACATTTGTCAGGCAAAAAGTATATAAATAGGCAAAACTCTACCCCTTTTA 98
Db　67598　CTGCTTTATTGCAAAGTTAGTTAGTTAGATCAAAATTAGAAATCATCCACCACTACTCTAAAGG 67539

Qy　　99　TTTTAAATTAGATTTGGTGTCTGTGATTTGATGTCTGACGGGGACTGAGAGTAAATGGC 149
Db　67538　GTTTAAARGTGCCTATGAGTGAATAGACTTGAGAAGGGAGTTACAGTGTTGGC 67488

Query Match　　Score 37.4;　DB 9;

repeat_region　　31675. .31748
　　　　/rpt_family="L1"
repeat_region　　31749. .32053
　　　　/rpt_family="Alu"
repeat_region　　32054. .32636

repeat_region　　11301. .11450
　　　　/rpt_family="Alu"
repeat_region　　14573. .14710
　　　　/rpt_family="MIR"
misc_feature　　　14794. .15303
　　　　/rpt_family="L2"
misc_feature　　　14795. .15213
　　　　/note="similar to　EST H10802 (NID:g875622) ym04a11.r1"
repeat_region　　14939. .15077
　　　　/note="similar to　EST R51446 (NID:g813348) yg72h06.r1"
repeat_region　　15225. .15399
　　　　/rpt_family="MER1_type"
misc_feature　　　15452. .15722
　　　　/note="similar to　EST AL119629 (NID:g5925528)"
repeat_region　　15761. .15898
　　　　/rpt_family="MER1_type"
repeat_region　　15899. .16204
　　　　/rpt_family="Alu"
repeat_region　　16299. .16647
　　　　/rpt_family="Alu"
misc_feature　　　16398. .18721
　　　　/rpt_family="MER2_type"
repeat_region　　19001. .19248
　　　　/rpt_family="Alu"
repeat_region　　19628. .19813
　　　　/rpt_family="L1"
repeat_region　　19888. .20009
　　　　/rpt_family="L2"
repeat_region　　21112. .21200
　　　　/rpt_family="Alu"
repeat_region　　21210. .21512
　　　　/rpt_family="L2"
repeat_region　　21731. .21879
　　　　/rpt_family="Alu"
repeat_region　　21880. .22750
　　　　/rpt_family="L2"
misc_feature　　　23274. .23701
　　　　/rpt_family="ERV1"
repeat_region　　24064. .24171
　　　　/rpt_family="L2"
repeat_region　　24276. .25031
　　　　/rpt_family="MIR"
repeat_region　　25090. .25405
　　　　/rpt_family="L2"
repeat_region　　25416. .25756
　　　　/rpt_family="Alu"
misc_feature　　　25638. .26008
　　　　/note="similar to　EST AW593216 (NID:g7280474) hg11c01.x1"
misc_feature　　　25940. .26422
　　　　/note="similar to　EST AW892899 (NID:g8057104)"
repeat_region　　26585. .26626
　　　　/rpt_family="MaLR"
misc_feature　　　26810. .26904
　　　　/rpt_family="MIR"
misc_feature　　　27656. .28032
　　　　/note="similar to　EST AW593216 (NID:g7280474) hg11c01.x1"
misc_feature　　　27906. .27955
　　　　/note="similar to　EST AA961265 (NID:g3127282) on96a03.s1"
misc_feature　　　28132. .28330
　　　　/note="similar to　EST AW593216 (NID:g7280474) hg11c01.x1"
repeat_region　　28366. .28563
　　　　/rpt_family="L1"
repeat_region　　28559. .28952
　　　　/rpt_family="L1"
misc_feature　　　28685. .29040
　　　　/note="similar to　EST AA082647 (NID:g1624764) zn24a06.r1"
repeat_region　　30275. .30574
　　　　/rpt_family="Alu"
repeat_region　　30687. .31002
　　　　/rpt_family="MaLR"
repeat_region　　31148. .31587
　　　　/rpt_family="L1"
```

```
BASE COUNT      53622 a    33686 c    32247 g    52622 t
ORIGIN
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="RP11-96C4"
                /clone_lib="RPCI-11.1"

Query Match         7.4%;   Score 37.2;   DB 9;   Length 172177;
Best Local Similarity  50.0%;   Pred. No. 5.5;
Matches  93;   Conservative   0;   Mismatches   93;   Indels    0;   Gaps   0;

Qy     42 CTTTATGGGAACATTTGTCAGGCAAAAGTATAATAAATGGCAAACCTCACGCCTTTTATTT 101
          ||| | | |||||||| ||||||| | ||| | || || |  |||| |||| ||||||||
Db   1837 CTGTTAAGTAACACTTTGGTAAACACTTGGTCGCAATATTACACTTTTACACATCCTATTT 1896

Qy    102 TAAATTAGATTTGGTGTGATTTGATGCTGACGGGAGTGAGATAATGGCCTTATCCTGCTG 161
          |  || |||| | ||||||| | |||| |||| |  | ||  |||||| ||| ||||||
Db   1897 GACTCAGATCTTCTGTGTGGAATTCTTAGTGACCTTTCTGAGGGGTGAGGTTTGGTGTGCTG 1956

Qy    162 CAGGCTGTGCTGAGGATGGCCTGGTCTGCCACCCTCCTGCAGTAGCATTTTGCATGTGTA 221
          ||| ||||| || |  ||| |  | | |||| ||||| |||||| |  || ||||||||
Db   1957 CATGCTTTGTCAAGAAAAGAAGTGGGTTACCTTATTGCTCAGGTTCACATTCGCATTTCTA 2016

Qy    222 ACAGGG 227
          ||||||
Db   2017 CCGAGG 2022

RESULT 12
LOCUS       AC010159
ACCESSION   AC010159   233721 bp   DNA   linear   HTG  06-FEB-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-326B1, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
VERSION     AC010159.9  GI:13940587
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN;
            HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 233721)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 233721)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-1999) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On May 4, 2001 this sequence version replaced gi:13605962.
            --------------- Genome Center
            Center: Genome Therapeutics Corporation
            Center code: GTC
            Web site: http://www.genomecorp.com/
            Contact: gtc-seqcenter@genomecorp.com
            --------------- Project Information
            Center project name: hg014
            --------------- Summary Statistics
            Sequencing vector: N/A
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly Program: Phrap; version 990315
            Consensus quality: 221102 bases at least Q40
            Consensus quality: 223407 bases at least Q30
            Consensus quality: 224757 bases at least Q20
            Insert size: 232071; sum-of-contigs
            Quality coverage: 6.7x in Q20 bases; sum-of-contigs
            --------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 18 contigs. The true order of the pieces is
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
```

FEATURES   Location/Qualifiers
     source   1..233721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-326B1"
                     /clone_lib="RPCI-11"
     misc_feature   1..1092
                     /note="assembly_name:Contig1"
     misc_feature   1193..2327
                     /note="assembly_name:Contig97"
     misc_feature   2428..3525
                     /note="assembly_name:Contig9"
     misc_feature   3626..5063
                     /note="assembly_name:Contig15"
     misc_feature   5164..6410
                     /note="assembly_name:Contig32"
     misc_feature   6511..7710
                     /note="assembly_name:Contig34"
     misc_feature   7811..8891
                     /note="assembly_name:Contig39"
     misc_feature   8992..10718
                     /note="assembly_name:Contig41"
     misc_feature   10819..11857
                     /note="assembly_name:Contig42"
     misc_feature   11958..13382
                     /note="assembly_name:Contig49"
     misc_feature   13483..15605
                     /note="assembly_name:Contig50"
     misc_feature   15706..17861
                     /note="assembly_name:Contig54"
     misc_feature   17962..20319
                     /note="assembly_name:Contig55"

```
1093 : gap of 100 bp in length
1092 : contig of 1092 bp in length
1192 : gap of unknown length
2327 : contig of 1135 bp in length
2427 : gap of unknown length
3525 : contig of 1098 bp in length
3525 : gap of unknown length
3625 : contig of 1438 bp in length
3626 : gap of unknown length
5063 : contig of 1438 bp in length
5164 : gap of unknown length
6410 : contig of 1247 bp in length
6510 : gap of unknown length
7710 : contig of 1200 bp in length
7810 : gap of unknown length
8891 : contig of 1081 bp in length
8991 : gap of unknown length
10718 : contig of 1727 bp in length
10818 : gap of unknown length
11857 : contig of 1039 bp in length
11957 : gap of unknown length
13382 : contig of 1425 bp in length
13482 : gap of unknown length
15605 : contig of 2123 bp in length
15705 : gap of unknown length
17861 : contig of 2156 bp in length
17961 : gap of unknown length
20319 : contig of 2358 bp in length
20419 : gap of unknown length
20320 : contig of 3811 bp in length
24230 : gap of unknown length
24231 : contig of 32994 bp in length
24331 : gap of unknown length
57325 : contig of 36980 bp in length
57425 : gap of unknown length
94404 : contig of unknown length
94504 : gap of unknown length
94505 : contig of 45566 bp in length
140071 : gap of unknown length
140070 : contig of unknown length
140170 : gap of unknown length
140171 : contig of 93551 bp in length
233721 : contig of 93551 bp in length
```

```
misc_feature    20420..24230
                /note="assembly_name:Contig956"
misc_feature    24331..57324
                /note="assembly_name:Contig957
                clone_end:T7"
misc_feature    57425..94404
                /note="assembly_name:Contig958"
misc_feature    94505..140070
                /note="assembly_name:Contig959"
misc_feature    140171..233721
                /note="assembly_name:Contig960"
BASE COUNT      70671 a   45926 c   45132 g   70249 t   1743 others
ORIGIN

Query Match          7.48;    Score 37.2;   DB 2;   Length 233721;
Best Local Similarity  50.0%;   Pred. No. 5.7;
Matches     93;   Conservative   0;   Mismatches   93;   Indels   0;   Gaps   0;

Qy    42  CTTTATGGGGAACATTTGTCAGGCAAAAGTATAATAAATGGCAAAAACTCTACGGCCTTTTATTT 101
          |||  |||| ||||||||||  ||| |  ||| |   ||||| |||||| || |||| |||||||
Db  51151  CTGTTAAGTAACACTTTGGTAAACAACTTGGTGCAATATTACACTTTACACATCCTATTT 51216

Qy   102  TAAATTAGATTGGTGTGTGATTTGCATGCGCGGAGTGAGAGTAATGGCCTTATCCTGCTG 161
          |  |||||| ||||||| ||||||    |||| |||||||   || | ||||||| |||||
Db  51217  GACTCAGANCTTCTGTGAATTCTTAGTGACCTTTCTGAGGGGTGAGGTTTGGGTGTGCTG 51276

Qy   162  CAGGCTGTGTCGTGAGGATGGCCTGGTCGTGCGCCACCCTCCTCGAGTAGCATTTTGCATGTGTA 221
          || ||| || |||||||||||||||| ||| ||||  |||| ||  || |||||| ||  |||||
Db  51277  CATGCTTTTGTCAAGAAAGAAGTGGGTTTACCTTATTGCTCAGGTTCACANTCGGCANTTCTA 51336

Qy   222  ACAGGG 227
          ||||||
Db  51337  CCGAGG 51342
```

```
RESULT 13
AL391218/c
LOCUS       AL391218      173149 bp    DNA      linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-13G5, *** SEQUENCING IN
            PROGRESS ***, 20 unordered pieces.
ACCESSION   AL391218
VERSION     AL391218.9  GI:11322137
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173149)
AUTHORS     Burton,J.
TITLE       Direct Submission
JOURNAL     Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 23, 2000 this sequence version replaced gi:11022482.
            -------------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------------- Project Information
            Center project name: bA13G5
            -------------------- Summary Statistics
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 2% of reads Chemistry:
            Dye-terminator Big Dye; 98% of reads
            Assembly program: XGAP4; version 4.5
            Consensus quality: 163128 bases at least Q40
            Consensus quality: 167065 bases at least Q30
            Consensus quality: 169494 bases at least Q20
            Insert size: 171249; sum-of-contigs
            Insert size: 136066; 25.3% error:agarose-fp
            Quality coverage: 4.17x in Q20 bases; sum-of-contigs Quality
            coverage: 5.40x in Q20 bases; agarose-fp
```

```
*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 20 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*
*       1  2724: contig of 2724 bp in length
*       2725  2824: gap of  100 bp
*       2825  9154: contig of 6330 bp in length
*       9155  9254: gap of  100 bp
*       9255  12637: contig of 3383 bp in length
*      12638  12737: gap of  100 bp
*      12738  21028: contig of 8291 bp in length
*      21029  21128: gap of  100 bp
*      21129  29585: contig of 8457 bp in length
*      29586  29685: gap of  100 bp
*      29686  32087: contig of 2402 bp in length
*      32088  32187: gap of  100 bp
*      32188  45535: contig of 13348 bp in length
*      45536  45635: gap of  100 bp
*      45636  66816: contig of 21181 bp in length
*      66817  66916: gap of  100 bp
*      66917  88672: contig of 21756 bp in length
*      88673  88772: gap of  100 bp
*      88773  97718: contig of 8946 bp in length
*      97719  97818: gap of  100 bp
*      97819  102378: contig of 4560 bp in length
*     102379  102478: gap of  100 bp
*     102479  105072: contig of 2594 bp in length
*     105073  105172: gap of  100 bp
*     105173  108462: contig of 3290 bp in length
*     108463  108562: gap of  100 bp
*     108563  118871: contig of 10309 bp in length
*     118872  118971: gap of  100 bp
*     118972  125792: contig of 6821 bp in length
*     125793  125892: gap of  100 bp
*     125893  134508: contig of 8616 bp in length
*     134509  134608: gap of  100 bp
*     134609  139070: contig of 4462 bp in length
*     139071  139170: gap of  100 bp
*     139171  144122: contig of 4952 bp in length
*     144123  144222: gap of  100 bp
*     144223  153507: contig of 9285 bp in length
*     153508  153607: gap of  100 bp
*     153608  173149: contig of 19542 bp in length.
FEATURES             Location/Qualifiers
     source          1..173149
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-13G5"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..2724
                     /note="assembly_fragment:00474
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2825..9154
                     /note="assembly_fragment:00363
                     fragment_chain:1"
     misc_feature    9255..12637
                     /note="assembly_fragment:00235
                     fragment_chain:1"
     misc_feature    12738..21028
                     /note="assembly_fragment:00435
                     fragment_chain:1"
     misc_feature    21129..29585
                     /note="assembly_fragment:01281
                     fragment_chain:2"
     misc_feature    29686..32087
```

misc_feature   /note="assembly_fragment:00030
                fragment_chain:2"
       32188..45535   /note="assembly_fragment:00997
                fragment_chain:2"
misc_feature   45636..66816   /note="assembly_fragment:00097
                fragment_chain:3"
misc_feature   66917..88672   /note="assembly_fragment:01005
                fragment_chain:3"
misc_feature   88773..97718   /note="assembly_fragment:00100
                fragment_chain:4"
misc_feature   97819..102378   /note="assembly_fragment:01502
                fragment_chain:4"
       102479..105072   /note="assembly_fragment:00096"
misc_feature   105173..108462   /note="assembly_fragment:00097"
misc_feature   108563..118871   /note="assembly_fragment:00400"
misc_feature   118972..125792   /note="assembly_fragment:00400"
misc_feature   125893..134508   /note="assembly_fragment:00413"
misc_feature   134609..139070   /note="assembly_fragment:00537"
misc_feature   139171..144122   /note="assembly_fragment:00679"
misc_feature   144223..153507   /note="assembly_fragment:00914"
misc_feature   153608..173149   /note="assembly_fragment:01096"
       173149..   /note="assembly_fragment:00397
                clone_end:T7
                vector_side:right"

BASE COUNT   46551 a   39780 c   39716 g   45190 t   1912 others
ORIGIN

RESULT 14
Best Local Similarity 54.9%;   Pred. No. 6.3;
Query Match   7.4%;   Score 37;   DB 2;   Length 173149;
Matches 73; Conservative 0;   Mismatches 60;   Indels 0;   Gaps 0;

Qy   341 CTATGGCCAGTCCTCCCAGAGCTCTGCCAGCTGCACTTGGGGGTGGCAGTCTGTGTTG 400
            ||||||| ||||||||| || ||||||||||||||||||||| ||| ||| ||
Db 134474 CCAGGGACCAGGCTCCCACAGCAAGGACCAGCAGCTCCCAGCGGGAGTAAAATGAGGTGGGTG 134415

Qy   401 TCCTGCGTGATAACGGCGCGTGAAAGCCACCAACTGCTGCCCAAAAATCACCCAGCCGATT 460
            ||||| || |||||||||||||| | | ||| ||| | || || | ||
Db 134414 TCCCCGCTGCACAGGCGCGGGACACAGCGTCACCTGTTGTGTGAGACCTGCGTGCCTTGT 134355

Qy   461 GGGGGTTCCCAT 473
            ||||| ||||
Db 134354 CGGGGTCCACCAT 134342

LOCUS   AL590128   204515 bp   DNA   linear   HTG 20-DEC-2001
DEFINITION   Homo sapiens chromosome 1 clone RP11-312B8, *** SEQUENCING IN
          PROGRESS ***, in ordered pieces.
ACCESSION   AL590128 AC036238
VERSION   AL590128.7 GI:17221201
KEYWORDS   HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

FEATURES   Location/Qualifiers
     source   1..204515
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-312B8"
            /clone_lib="RPCI-11.2"
     misc_feature   1..204515
            /note="assembly_fragment:03105"

BASE COUNT   51937 a   44093 c   47789 g   60696 t
ORIGIN

RESULT 15
Best Local Similarity 54.9%;   Pred. No. 6.5;
Query Match   7.4%;   Score 37;   DB 2;   Length 204515;
Matches 73; Conservative 0;   Mismatches 60;   Indels 0;   Gaps 0;

Qy   341 CTATGGCCAGTCCTCCCAGAGCCTCTGCCAGCGTGCCACTTGGGGGTGGGACAGTCTCGTGCTTG 400
            ||||||| ||||||||| || ||||||||||||||||||||| ||| ||| ||
Db 185819 CCAGGGACCAGGCTCCCACAGCAAGGACCAGCAGCTCCCAGCGGGAGTAAAATGAGGTGGGTG 185760

Qy   401 TCCTGCGTGATAACGGCGCGTGAAAGCCAGCCAACTGCTGCCCAAAAATCACCCAGCCGATT 460
            ||||| || |||||||||||||| | | ||| ||| | || || | ||
Db 185759 TCCCCGCTGCACAGGCGCGGGACACAGCGTCACCTGTTGTGTGAGACCTGCGTGCCTTGT 185700

Qy   461 GGGGGTTCCCAT 473
            ||||| ||||
Db 185699 CGGGGTCCACCAT 185687

LOCUS   AP002402   146691 bp   DNA   linear   HTG 31-MAY-2000
DEFINITION   Homo sapiens chromosome 18 clone RP11-731I21 map 18q12, WORKING
          DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION   AP002402
VERSION   AP002402.1 GI:8131666
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens DNA, clone:RP11-731I21.
ORGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 146691)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 146,691 genomic DNA of 18q12
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 146691)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ. 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp, Tel:81-42-778-9923,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     -------------- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
            Center project name: Humdraft18
            Center clone name: RP11-731I21
            -------------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Assembly quality: 134605 bases at least Q40
            Consensus quality: 141033 bases at least Q40
            Consensus quality: 143879 bases at least Q30
            Consensus quality: 143879 bases at least Q20
            Insert size: 144991; sum-of-contigs
            Consensus quality: 144991; sum-of-contigs
            Quality coverage: 4.32x in Q20 bases; sum-of-contigs
            --------------
            NOTE: This is a 'working draft' sequence. It currently consists of
            18 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs of N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved.
```

```
          *      1     33964: contig of  33964 bp in length
          *  33965     34064: gap of       100 bp
          *  34065     50706: contig of  16642 bp in length
          *  50707     50806: gap of       100 bp
          *  50807     69999: contig of  19193 bp in length
          *  70000     70099: gap of       100 bp
          *  70100     83976: contig of  13877 bp in length
          *  83977     84076: gap of       100 bp
          *  84077     96001: contig of  11925 bp in length
          *  96002     96101: gap of       100 bp
          *  96102    106846: contig of  10745 bp in length
          * 106847    106946: gap of       100 bp
          * 106947    115930: contig of   8984 bp in length
          * 115931    116030: gap of       100 bp
          * 116031    121613: contig of   5583 bp in length
          * 121614    121713: gap of       100 bp
          * 121714    125910: contig of   4197 bp in length
          * 125911    126010: gap of       100 bp
          * 126011    129235: contig of   3225 bp in length
          * 129236    129335: gap of       100 bp
          * 129336    131878: contig of   2543 bp in length
          * 131879    131978: gap of       100 bp
          * 131979    134377: contig of   2399 bp in length
          * 134378    134477: gap of       100 bp
          * 134478    137338: contig of   2861 bp in length
          * 137339    137438: gap of       100 bp
          * 137439    139693: contig of   2255 bp in length
          * 139694    139793: gap of       100 bp
          * 139794    142178: contig of   2385 bp in length
          * 142179    142278: gap of       100 bp
          * 142279    143695: contig of   1417 bp in length
          * 143696    143795: gap of       100 bp
          * 143796    145334: contig of   1539 bp in length
          * 145335    145434: gap of       100 bp
          * 145435    146691: contig of   1257 bp in length.
FEATURES             Location/Qualifiers
     source          1..146691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q12"
                     /clone="RP11-731I21"
     misc_feature    1..33964
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    34065..50706
                     /note="assembly_fragment"
     misc_feature    50807..69999
                     /note="assembly_fragment"
     misc_feature    70100..83976
                     /note="assembly_fragment"
     misc_feature    84077..96001
                     /note="assembly_fragment"
     misc_feature    96102..106846
                     /note="assembly_fragment"
     misc_feature    106947..115930
                     /note="assembly_fragment"
     misc_feature    116031..121613
                     /note="assembly_fragment"
     misc_feature    121714..125910
                     /note="assembly_fragment"
     misc_feature    126011..129235
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    129336..131878
                     /note="assembly_fragment"
     misc_feature    131979..134377
                     /note="assembly_fragment"
     misc_feature    134478..137338
                     /note="assembly_fragment"
     misc_feature    137439..139693
                     /note="assembly_fragment"
     misc_feature    139794..142178
                     /note="assembly_fragment"
     misc_feature    142279..143695
                     /note="assembly_fragment"
     misc_feature    143796..145334
                     /note="assembly_fragment"
     misc_feature    145435..146691
                     /note="assembly_fragment"
BASE COUNT     45915 a   27527 c   27197 g   44352 t   1700 others
ORIGIN
```

```
Query Match        7.3%;    Score 36.6;    DB 2;    Length 146691;
Best Local Similarity    60.6%;    Pred. No. 8.2;
Matches   60;  Conservative   0;  Mismatches   39;  Indels   0;  Gaps   0;

Qy     13 ATCTATTTTGGACTATGCCTTGCATACAGCTTTATGGGAACACATTTGTCAGGCAAAAGTAT 72
          || | |||| || | ||||  || |||| ||||||| ||| || ||||| || || ||
Db  98384 ATATGCTTGGGACAAATTCATTTGATAAAGCTTAATAGAAAAGTTATATGTATCACAATAACAT 98325

Qy     73 AATAATGGCAAACTCTACGCCCTTTTATTTTAAATTAGAT 111
          ||||| ||| |||  ||| | ||
Db  98324 CATAATTGCAGATTCTAATTTAATTATATGTGATTTGTACAT 98286
```

Search completed: October 10, 2002, 18:16:36
Job time : 1353.87 secs

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Title: US-09-489-101A-13_COPY_1_500
Perfect score: 500
Sequence: 1 gcccgaaaccggaagtgag..........gcgactttgctcaaggactcc 500

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*

Search time 715.867 Seconds
(without alignments)
14616.225 Million cell updates/sec

```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_Om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_hgo_inv:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.		Score	% Query Match	Length		DB ID	Description
1		500	100.0	500	6	AX201592	AX201592 Sequence
2		500	100.0	4877	9	AB023180	Homo sapi
3	c	221.8	44.4	35848	2	AC011553	Homo sapi
4	c	114.4	22.9	6153	9	AC093067	Homo sapi
5	c	72.4	14.5	221824	2	AC069214	Homo sapi
6	c	71.8	14.4	115715	2	AC079411	Homo sapi
7	c	70.2	14.0	183920	2	AC023814	Homo sapi
8	c	69.6	13.9	197669	2	AC084064	Homo sapi
9	c	69.4	13.9	52804	2	AC091024	Homo sapi
10	c	69.2	13.8	209887	.	AC079424	Mus muscu
11	c	69.2	13.8	303091	2	AC084799	Mus muscu
12	c	69	13.8	66507	2	AC096161	Rattus no
13	c	69	13.8	158167	2	AC034105	Homo sapi
14	c	68	13.6	840	8	AL144464	Botrytis
15	c	68	13.6	143435	2	AC105671	Rattus no
16	c	67.8	13.6	78202	2	AC023212	Homo sapi
17	c	67.4	13.5	106950	2	AC034105	Homo sapi
18	c	67.4	13.5	161307	2	AC016474	Homo sapi
19	c	67.2	13.4	298166	2	AC084058	Homo sapi
20	c	67	13.4	56270	2	AC087563	Homo sapi
21	c	67	13.4	63822	2	AC102429	Mus muscu
22	c	66.8	13.4	72360	2	AC024277	Homo sapi
23	c	66.8	13.4	82897	2	AC022984	Homo sapi
24	c	66.8	13.4	143044	2	AC022595	Homo sapi
25	c	66.8	13.3	16274	2	AC094826	Rattus no
26	c	66.6	13.3	141892	2	AC110453	Rattus no
27	c	66.6	13.3	300695	2	AC023197	Mus muscu
28	c	66.2	13.2	75628	2	AC021793	Homo sapi
29	c	66	13.2	220469	2	AC074307	Mus muscu
30	c	66	13.2	33058	2	AC100211	Mus muscu
31	c	65.8	13.2	137878	2	AP003300	Oryza sat
32	c	65.4	13.1	461181	2	AC105855	Rattus no
33	c	65.2	13.1	69515	2	AC099971	Mus muscu
34	c	65.2	13.0	105965	2	AC108986	Rattus no
35	c	65.2	13.0	181988	2	AC090552	Homo sapi
36	c	65	13.0	56870	2	CEX39B6_3	Continuation (4 of
37	c	65	13.0	73848	2	AC022556	Homo sapi
38	c	65	13.0	110737	2	AC011105	Homo sapi
39	c	65	13.0	249262	2	AC079430	Homo sapi
40	c	65	13.0	79355	2	AC021280	Homo sapi
41	c	64.8	13.0	97353	2	AC095846	Rattus no
42	c	64.8	13.0	265537	2	AC087228	Rattus no
43	c	64.8	13.0	97353	2	AC087228	Mus muscu
44	c	64.8	12.9	64651	2	AC101572	Mus muscu
45	c	64.6	12.9	204889	2	AL627343	Mus muscu

RESULT 1
AX201592

```
LOCUS       AX201592      4877 bp    DNA    linear    PAT 30-AUG-2001
DEFINITION  Sequence 13 from Patent WO0153349.
ACCESSION   AX201592
VERSION     AX201592.1  GI:15391441
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4877)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE     Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 13 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES             Location/Qualifiers
     source          1..4877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

ALIGNMENTS

```
BASE COUNT      867 a   1682 c   1570 g    758 t
ORIGIN

  Query Match             100.0%;  Score 500;  DB 6;  Length 4877;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-67;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCCCGAAACCCGGAAGTGAGCGGCGGCAGCTGCGAGGCTCGGAGAAACAGGCGCCGCGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCCGAAACCCGGAAGTGAGCGGCGGCAGCTGCGAGGCTCGGAGAAACAGGCGCCGCGGG 60

Qy         61 CTCCGCGCCCGGCCGGACCCGGGCCCGAGATCATGATGCTGCCGCCACCGCCGCCACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCGCGCCCGGCCGGACCCGGGCCCGAGATCATGATGCTGCCGCCACCGCCGCCACCAC 120

Qy        121 GGAGCGAGAAGCCCAGATAGACGCCCCGGCGGCCCCGGGTCCTGGAGTCCCGCCGCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGAGCGAGAAGCCCAGATAGACGCCCCGGCGGCCCCGGGTCCTGGAGTCCCGCCGCCTGC 180

Qy        181 TGCCCGGCCGAGGACCCCACCCCGCCTGCCGCCCGATGCTTGCAGTGGGGCCCGCCATGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGCCCGGCCGAGGACCCCACCCCGCCTGCCGCCCGATGCTTGCAGTGGGGCCCGCCATGG 240

Qy        241 ACAGGGATTACCCGCAGCATGAACCCCCGCCGGCGGGCAGCCTCCTGTACAGCCCGCCGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACAGGGATTACCCGCAGCATGAACCCCCGCCGGCGGGCAGCCTCCTGTACAGCCCGCCGC 300

Qy        301 CCCTGCAGAGCGCCATGCTGCACTGCCCCTACTGGAACACCTTCTCGCTGCCGCCATACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCTGCAGAGCGCCATGCTGCACTGCCCCTACTGGAACACCTTCTCGCTGCCGCCATACC 360

Qy        361 CTGCCTTCTCCAGCGACAGCCGCCCGTTCATGAGCTCCGCCTCCTTCCTCGGCAGCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCCTTCTCCAGCGACAGCCGCCCGTTCATGAGCTCCGCCTCCTTCCTCGGCAGCCAGC 420

Qy        421 CCTGCCCAGACACCAGCTATGCCCCCGTGGCCACCGCCTCCAGCTTGCCACCAAAGACCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGCCCAGACACCAGCTATGCCCCCGTGGCCACCGCCTCCAGCTTGCCACCAAAGACCT 480

Qy        481 GCGACTTTGCTCAGGACTCC 500
              ||||||||||||||||||||
Db        481 GCGACTTTGCTCAGGACTCC 500

RESULT 2
AB023180
LOCUS       AB023180                4877 bp    mRNA    linear   PRI 16-JUN-1999
DEFINITION  Homo sapiens mRNA for KIAA0963 protein, complete cds.
ACCESSION   AB023180
VERSION     AB023180.1  GI:4589569
KEYWORDS    .
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
            SK plus clone:hj05936.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosawa,M.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 6 (1), 63-70 (1999)
  MEDLINE   99246063
REFERENCE   2  (bases 1 to 4877)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..4877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hj05936"
                     /sex="male"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
     gene            216..4316
                     /gene="KIAA0963"
     CDS             216..4316
                     /gene="KIAA0963"
                     /codon_start=1
                     /product="KIAA0963 protein"
                     /protein_id="BAA76807.1"
                     /db_xref="GI:4589570"
                     /translation="MLAVGPAMDRDYPQHEPPPAGSLLYSPPPLQSAMLHCPYWNTFS
                     LPPYPAFSSDSRPFMSSASFLGSQPCPDTSYAPVATASSLPPKTCDFAQDSSYFEDFS
                     NISIFSSSVDSLSDIVDTPDFLPADSLNQVSTIWDDNPAPSTHDKLFQLSRPFAGFED
                     FLPSHSTPLLVSYQEQSVQSQPEEEDEAEEEEAEELGHTETYADYVPSKSKIGKQHPD
                     RVVETSTLSSVPPPDITYTLALPSDSGALSALQLEAITYACQQHEVLLPSGQRAGFLI
                     GDGAGVGKGRTVAGVILENHLRGRKKALWFSVSNDLKYDAERDLRDIEATGIAVHALS
                     KIKYGDTTTSEGVLFATYSALIGESQAGGQHRTRLRQILDWCGEAFEGVIVFDECHKA
                     KNAGSTKMGKAVLDLQNKLPLARVVYASATGASEPRNMIYMSRLGIWGEGTPFRNFEE
                     FLHAIEKRGVGAMEIVAMDMKVSGMYIARQLSFSGVTFRIEEIPLAPAFECVYNRAAL
                     LWAEALNVFQQAADWIGLESRKSLWGQFWSAHQRFFKYLCIAAKVRRLVELAREELAR
                     DKCVVIGLQSTGEARTREVLGENDGHLNCFVSAAEGVFLSLIQKHFPSTKRKRDRGAG
                     SKRKRRPRGRGAKAPRLACETAGVIRISDDSSTESDPGLDSDFNSSPESLVDDDVVIV
                     DAVGLPSDDRGSLCLLQRDPHGPGVLERVERLKQDLLDKVRRLGRELPVNTLDELIDQ
                     LGGPQRVAEMTGRKGRVVSRPDGTVAFESRAEQGLSIDHVNLREKQRFMSGEKLVAII
                     SEASSSGVSLQADRRVQNQRRRVHMTLELPWSADRAIQQFGRTHRSNQVSAPEYVFLI
                     SELAGERRFASIVAKRLESLGALTHGDRRATESRDLSKYNFENKYGTRALHCVLTTIL
                     SQTENKVPVPQGYPGGVPTFFRDMKQGLLSVGIGGRESRNGCLDVEKDCSITKFLNRI
                     LGLEVHKQNALFQYFSDTFDHLIEMDKREGKYDMGILDLAPGIEEIYEESQQVFLAPG
                     HPQDGQVVFYKISVDRGLKWEDAFAKSLALTGPYDGFYLSYKVRGNKPSCLLAEQNRG
                     QFFTVYKPNIGRQSQLEALDSLRRKFHRVTAEEAKEPWESGYALSLTHCSHSAWNRHC
                     RLAQEGKDCLQGLRLRHHYMLCGALLRVWGRIAAVMADVSSSSYLQIVRLKTKDRKKQ
                     VGIKIPEGCVRRVLQELRLMDADVKRRQAPALGCPAPPAPRPLALPCGPGEVLDLTYS
                     PPAEAFPPPPHFSFPAPLSLDAGPGVVPLGTPDAQADPAALAHQGCDINFKEVLEDML
                     RSLHAGPPSEGALGEGAGAGGAAGGGPERQSVIQFSPPFPGAQAPL"
BASE COUNT      867 a   1682 c   1570 g    758 t
ORIGIN

  Query Match             100.0%;  Score 500;  DB 9;  Length 4877;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-67;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCCCGAAACCCGGAAGTGAGCGGCGGCAGCTGCGAGGCTCGGAGAAACAGGCGCCGCGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCCCGAAACCCGGAAGTGAGCGGCGGCAGCTGCGAGGCTCGGAGAAACAGGCGCCGCGGG 60

Qy         61 CTCCGCGCCCGGCCGGACCCGGGCCCGAGATCATGATGCTGCCGCCACCGCCGCCACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCGCGCCCGGCCGGACCCGGGCCCGAGATCATGATGCTGCCGCCACCGCCGCCACCAC 120

Qy        121 GGAGCGAGAAGCCCAGATAGACGCCCCGGCGGCCCCGGGTCCTGGAGTCCCGCCGCCTGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGAGCGAGAAGCCCAGATAGACGCCCCGGCGGCCCCGGGTCCTGGAGTCCCGCCGCCTGC 180

Qy        181 TGCCCGGCCGAGGACCCCACCCCGCCTGCCGCCCGATGCTTGCAGTGGGGCCCGCCATGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGCCCGGCCGAGGACCCCACCCCGCCTGCCGCCCGATGCTTGCAGTGGGGCCCGCCATGG 240

Qy        241 ACAGGGATTACCCGCAGCATGAACCCCCGCCGGCGGGCAGCCTCCTGTACAGCCCGCCGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACAGGGATTACCCGCAGCATGAACCCCCGCCGGCGGGCAGCCTCCTGTACAGCCCGCCGC 300

Qy        301 CCCTGCAGAGCGCCATGCTGCACTGCCCCTACTGGAACACCTTCTCGCTGCCGCCATACC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCTGCAGAGCGCCATGCTGCACTGCCCCTACTGGAACACCTTCTCGCTGCCGCCATACC 360

Qy        361 CTGCCTTCTCCAGCGACAGCCGCCCGTTCATGAGCTCCGCCTCCTTCCTCGGCAGCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCCTTCTCCAGCGACAGCCGCCCGTTCATGAGCTCCGCCTCCTTCCTCGGCAGCCAGC 420
```

RESULT 3
Qy   421 CCTGCCCAGACACCAGCTATGCCCCGTGGCCCCACCGCCTTCCAGTTGCCACCAAAGACCT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCTGCCCAGACACCAGCTATGCCCCGTGGCCCCACCGCCTTCCAGTTGCCACCAAAGACCT 480

Qy   481 GCGACTTTGCTCAGGACTCC 500
           ||||||||||||||||||||
Db   481 GCGACTTTGCTCAGGACTCC 500

LOCUS   AC011553/c   35848 bp   DNA   linear   HTG 23-APR-2001
DEFINITION Homo sapiens chromosome 19 clone LLNLR-277D11, WORKING DRAFT
            SEQUENCE, 3 ordered pieces.
ACCESSION   AC011553
VERSION   AC011553.3 GI:7711540
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 35848)
AUTHORS   DOE Joint Genome Institute.
TITLE   Sequencing of Human Chromosome 19
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 35848)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On May 6, 2000 this sequence version replaced gi:7690187.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 26813, R31543
            Center clone name: LLNLR_277D11

            Summary Statistics
            Consensus quality: 35124 bases at least Q40
            Consensus quality: 35398 bases at least Q30
            Consensus quality: 35534 bases at least Q20
            Estimated insert size: 32650; agarose-fp estimation
            Estimated insert size: 35748; sum-of-contigs estimation
            Quality coverage: 10.55 in Q20 bases; agarose-fp estimation
            Quality coverage: 9.64 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 3 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces
            is believed to be correct as given, however the sizes
            of the gaps between them are based on estimates that have
            provided by the submitter.
            This sequence will be replaced
            by the finished sequence as soon as it is available and
            the accession number will be preserved.
             1   18613: contig of 18613 bp in length
            18614 18713: gap of unknown length
            18714 19277: contig of 564 bp in length
            19278 19377: gap of unknown length
            19378 35848: contig of 16471 bp in length.
FEATURES   Location/Qualifiers
     source   1..35848
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone_lib="Lawrence Livermore human cosmid library LLNLR"
                     /clone="LLNLR-277D11"
BASE COUNT   7758 a   11001 c   10770 g   6119 t   200 others
ORIGIN

Query Match   44.4%; Score 221.8; DB 2; Length 35848;
Best Local Similarity   91.4%; Pred. No. 2.3e-25;
Matches 235; Conservative   0; Mismatches   22; Indels   0; Gaps   0;

Qy   73 CCGGACCCGGGCGCCGAGATCATGATACTGCGGCCACCGGCCGCCACCGAGGACGGAGAGC 132
           || ||| | | ||||| |||||||||||||||||| ||||||| ||||| |||
Db   5256 CCCGTCTCCCCTCTGCAGATCATGATGAGGCTGCGGCCGCCACCGGCCGCCACCGGGAAGC 5197

Qy   133 CCAGATAGACGACCCCCGGCGGCCCCGGCGGCCGGGTCCTTGGAGTCCCCGCCGCCCCGAG 192
           |||| || || |||||||| ||||||||||||||| |||||||||||||||||||||||
Db   5196 CCAGATAGACGCCCCCCGGCGGCCCCGGCGGCCGGGTCCTTGGAGTCCCCGCCGCCCCGAG 5137

Qy   193 GACCCCCACCCCGGCCTGCGCGGCCCGATGGCTTGCAGTGTGGCCGCCGCCATGGACAGGA 252
           ||| ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   5136 GACCCCACCCCGGCCTGCGCGGCCCGATGGCTTGCAGTGGGGCCGCCGCCATGGACAGGA 5077

Qy   253 CGCAGCATGAAACCCCCGCGGGGGCAGCCTCCTGTACAGCCCCGCCCCGGCCCTGCAGAGCG 312
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5076 CGCAGCATGAAACCCCCGCGGGGGCAGCCTCCTGTACAGCCCCGCCCCGGCCCTGCAGAGCG 5017

Qy   313 CCATGCTGCACTGCCCC 329
           | |||| ||||||
Db   5016 GCCCCACCCCGGCCCC 5000

RESULT 4
LOCUS   AC093067/c   6153 bp   DNA   linear   PRI 24-OCT-2001
DEFINITION Homo sapiens chromosome 19 clone LLNLR-240D7, complete sequence.
ACCESSION   AC093067
VERSION   AC093067.2 GI:16356878
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 6153)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 6153)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (09-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 6153)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (24-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   On Oct 24, 2001 this sequence version replaced gi:15144408.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 98.4% of Sequence;
            Estimated Total Number of Errors is 0.
            NOTE: This sequence is not the entire sequence of the clone. It
            is sequence generated to span the gap between AC005390 and
            AC011553. The overlap with AC005390 is 1172bp and the overlap with
            AC011553 is 1062bp. The sequence was finished by the Stanford
            Human Genome Center and Los Alamos National Laboratory.
FEATURES   Location/Qualifiers
     source   1..6153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="LLNLR-240D7"
BASE COUNT   1346 a   2032 c   1887 g   888 t
ORIGIN

Query Match   22.9%; Score 114.4; DB 9; Length 6153;
Best Local Similarity   99.1%; Pred. No. 8.7e-09;

RESULT 5

```
Qy     380 CCGGCCGTTCATGAGCTCCGCCGCTTCCTTCCTCGCCCCTTCCCCAGACACACAGCTA 439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    3350 CAGCCCGTTCATGAGCTCCGCCGCTTCCTTCCTCGCCCCTTCCCCAGACACACAGCTA 3291

Qy     440 TGCCCCCGTGGCCACCCCCTTCCAGCTTGCCACCAAAGACCTTGCCTCAGG 495
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3290 TGCCCCCGTGGCCACCCCCTCCAGCTTGCCACCAAAGACCTTGCCTCAGG 3235
```

Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

LOCUS   AC069214   221852 bp   DNA   linear   HTG 15-OCT-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-665J20, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC069214
VERSION   AC069214.7 GI:14547423
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 221852)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Brieva,M., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,K., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,J., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,S.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 221852)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jun 25, 2001 this sequence version replaced gi:12831281.
            --------------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------------- Project Information
            Center project name: HAVI
            Center clone name: RP11-665J20
            --------------------- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; Version 0.990329
            Consensus quality: 203531 bases at least 040
            Consensus quality: 214587 bases at least 030
            Consensus quality: 218777 bases at least Q20
            Estimated insert size: 217968; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 6.2x in Q20 bases; sum-of-contigs estimation
            ----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1: contig of 40587 bp in length
            *   40588: gap of unknown length
            *   40587: contig of 40587 bp in length
            *   40687: gap of unknown length
            *   40688: contig of 28643 bp in length
            *   69330: gap of unknown length
            *   69331: contig of 28979 bp in length
            *   69430: gap of unknown length
            *   69431: contig of 28979 bp in length
            *   98409: gap of unknown length
            *   98410: contig of 27234 bp in length
            *   98509: gap of unknown length
            *   125743: contig of 27234 bp in length
            *   125744: gap of unknown length
            *   125843: contig of 23814 bp in length
            *   125844: gap of unknown length
            *   149657: contig of 23814 bp in length
            *   149658: gap of unknown length
            *   149757: contig of 18274 bp in length
            *   149758: gap of unknown length
            *   168031: contig of 18274 bp in length
            *   168032: gap of unknown length
            *   168131: contig of 9054 bp in length
            *   168132: gap of unknown length
            *   186097: contig of 17966 bp in length
            *   186098: gap of unknown length
            *   186197: contig of 9054 bp in length
            *   186198: gap of unknown length
            *   195251: contig of 9564 bp in length
            *   195252: gap of unknown length
            *   195351: contig of 5752 bp in length
            *   195352: gap of unknown length
            *   204915: contig of 9564 bp in length
            *   204916: gap of unknown length
            *   205015: contig of 5752 bp in length
            *   205016: gap of unknown length
            *   210767: contig of 3526 bp in length
            *   210768: gap of unknown length
            *   210867: contig of 3526 bp in length
            *   210868: gap of unknown length
            *   214393: contig of 2369 bp in length
            *   214394: gap of unknown length
            *   214493: contig of 2369 bp in length
            *   214494: gap of unknown length
            *   216862: contig of 2369 bp in length
            *   216863: gap of unknown length
            *   216962: contig of 2786 bp in length
            *   216963: gap of unknown length
            *   219747: contig of 2786 bp in length
            *   219748: gap of unknown length
            *   219848: contig of 2004 bp in length
            *   219849: gap of unknown length
            *   221852: contig of 2004 bp in length.
FEATURES   Location/Qualifiers
     source   1..221852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-665J20"
BASE COUNT   62725 a 49354 c 47934 g 60526 t 1313 others
ORIGIN
```

Qy   48 CAGGCGCGCGGCGGGCTCGCCGCCGGGCCGGGAGGCCCGGGAGATCATGAGTCGCGGCCA 107
           ||||| |||| | |||||| |||||||||||||||| | ||||||||||||||||||
Db 217640 CCGGCGCCCCGGCGGGGGCCCCCCCCCCCCCGGGCCGACCCCTCCCCCCCCCCCCCC 217581

Query Match   14.5%; Score 72.4; DB 2; Length 221852;
Best Local Similarity   49.5%; Pred. No. 0.01;
Matches 216; Conservative 0; Mismatches 216; Indels 4; Gaps 1;

Qy   108 CCGCCGCGGCCACCACGGAGGGAGAAGCCCAGATAGAGCGCCCGGGGGCGGGCCCCGGGGTTCCTTGGAG 167
           ||||| | ||||||||||||||||||||||||| ||||||||||||||||| |||| |||||

Db 217580 CGCCCCCGCCCTTGCCTCGCCGGGGGGGGGGGGGGGGGC 217521

Qy 168 TCCCCGCCTGCTGCCCACGAGGAGACCCACCACCGCCGGATGCGTTGCCAGT- 226
         |||||| ||| | ||| |||||||||| ||||||||||| | |||| |||
Db 217520 CCCCCGCCCATGGCACAGGGATTACCCGCGCCGGGGGGCCCCCCCCCCCC 217461

Qy 227 ---GGGGCCCCGGCCCCATGGACAGGGATTACCCGCAGCATGAAACCCCCGGGGCAGGCTT 283
         ||||||||| ||| |||||||||||||||||||||||||||||||||||||
Db 217460 CCCGGGGCCCCGGCCCCATGGACAGGGATTACCCGCGCCGGGGGGCCCCCCC 217401

Qy 284 CCTGTACAGCCCGCCCTGCAGAGCGCCCACTGCGCCCCTACTGGAACACCTT 343
         ||| ||||| || || || || || || || |||||| ||| ||||| ||
Db 217400 CCCCGGGGCCGGCCCCGGCCCCGGGCCCGGCCCCCCCCCCCCCCCCCCCC 217341

Qy 344 CTCGCTGCCCGCCCCATACCCTGCCTTCTCCAGGACAGCGCCGCCGGTTCATGAGCGTCCGGCCCTC 403
         |||| ||| ||| ||||||||| | || | ||| || ||||| || |||
Db 217340 CCCCGGGGCCCCGCCCCGGCCCCGGGGCCCCGGGGCCCGGGGGGCC 217281

Qy 404 CTTCCTCGGCCAGCCAGCCCGCCCAGACACCAGCTATGCCCCCCCCGGGGCCACCGCCGCCTCCAG 463
         ||| ||| ||||| || ||| || || |||||| ||| || ||| |||
Db 217280 CCCCGGGCCGGCCCCGGCCCCGCCCCCCGCCCCCCCCCCCCCCCCG 217221

Qy 464 CTTTGCCACCAAAGACC 479
         || ||||
Db 217220 GCCCCCCCCCCCGCCC 217205

RESULT 6
AC079411/c
LOCUS   AC079411   115715 bp   DNA   linear   HTG 19-SEP-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2050B12, WORKING DRAFT
            SEQUENCE, 2 ordered pieces.
ACCESSION   AC079411
VERSION   AC079411.3 GI:15667996
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 115715)
AUTHORS   DOE Joint Genome Institute.
TITLE   Sequencing of Human Chromosome 16
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 115715)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Sep 19, 2001 this sequence version replaced gi:13786419.
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *
            *
FEATURES   Location/Qualifiers
     source   1..115715
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="CTD-2050B12"
BASE COUNT   24685 a   25637 c   33266 t   223 others
ORIGIN
        1   40023   40022: contig of 40022 bp in length
        40123 40122: gap of unknown length
        115715: contig of 75593 bp in length

Query Match   14.4%; Score 71.8; DB 2; Length 115715;
Best Local Similarity 48.9%; Pred. No. 0.014;
Matches 207; Conservative 0; Mismatches 215; Indels 1; Gaps 1;

Qy   51 GCGCGCGCGGGGCTCCGCGCCCGGCGCGACCCGGGCCCGAGATCATGATGGCTGCCGCACCG 110
         ||||||||| || || || || |||||||| ||| ||| ||| ||||| |||
Db 39740 GCGCGCGCCCGCGCCCGCCCGCCCCCGCGCGGGCGCGCGCGCGCGCGCGC 39681

Qy 111 CCGCGCCACCACGGACGGACGAGGAAGCCCAGATAGGACGCCCCGGGCGCGCGGGGGGTCCTTGGAGTCC 170
         |||| ||| || || || || || |||||| || ||| || || |||
Db 39680 CCCCGCCCCGCCCGCGCGGGCCGCGGCGCCCCCGCGCGGGCGCGCCC 39621

Qy 171 -CGCCGCGCTGCTGCCCGCGGCGGAGGACCCACCCCGCCGGCGCTGCCGCCGCGGTGCTTGCAGTGGG 229
         ||||| || || || || || || |||||| || ||| || ||| |||
Db 39620 GCGGCCCCGGCCGCGCCCCGGGGGGCCCCCCGGNCCCCGGCGGCCGC 39561

Qy 230 GCGCGCCATGGACAGGGATTACCCGCCAGCCATGAAACCCCCGGGGCCGAGGCCAGGCTCGTGTA 289
         ||||| ||||||||||||||||| || || ||| ||| ||| ||| |||
Db 39560 GCGCGCCCGCGCGCGCCGCGCCGCCCGCGCGGNGCGCGCGCGCGCGCGGGC 39501

Qy 290 CAGCCCGCCCGCCCCCTTGCCAGAGCGCCCATGCTCGCACTGCCCCTTACTGGGAAACACCTTCTCGCT 349
         ||| || || || || || || || || ||| || ||| || || ||| |||
Db 39500 CCCNCCCCCCCCCCCGGGCGCGCCCCCGCGCGCGCGCCCCCGCGCGCGGCCC 39441

Qy 350 GCCGCCATACCCTGCCTTCTCCAGGGACACGCGCCCCGTTCATGAGCGTCGGCCGCCTCCTTCCT 409
         ||| || || || || || || || |||||| || ||| ||| ||| |||
Db 39440 CCCCCCCCCCCCCCCCGCCCCCCCCCCNCCCCCCCCCCCCCCCCCCC 39381

Qy 410 CGGCAGCCAGCCCTGCCCAGACACCAGCTATGCCCCCGTGGCCACCGCCCTCAGCGTTGCC 469
         ||| ||| || || || || || || || ||| || ||| ||| |||
Db 39380 CCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCC 39321

Qy 470 ACC 472
         |||
Db 39320 CCC 39318

RESULT 7
AC023814/c
LOCUS   AC023814   183920 bp   DNA   linear   HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2159J19, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC023814
VERSION   AC023814.4 GI:13786365
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 183920)
AUTHORS   DOE Joint Genome Institute.
TITLE   Sequencing of Human Chromosome 16
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 183920)
AUTHORS   DOE Joint Genome Institute.

TITLE   Direct Submission
JOURNAL   Submitted (18-FEB-2000) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Apr 25, 2001 this sequence version replaced gi:9099093.
          -----Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov

          Project Information
          Center Project Name: 0
          Center clone name: CITB-H1_2159J19

          Summary Statistics
          Consensus quality: 166828 bases at least Q40
          Consensus quality: 172844 bases at least Q30
          Consensus quality: 175535 bases at least Q20
          Estimated insert size: 173000; agarose-fp estimation
          Estimated insert size: 183020; sum-of-contigs estimation
          Quality coverage: 8.85 in Q20 bases; agarose-fp estimation
          Quality coverage: 8.36 in Q20 bases; sum-of-contigs estimation
          NOTE: This is a 'working draft' sequence. It currently
          consists of 10 contigs. The true order of the pieces
          is not known and their order in this sequence record is
          arbitrary. Gaps between the contigs are represented as
          runs of N, but the exact sizes of the gaps are unknown.
          This record will be updated with the finished sequence
          as soon as it is available and the accession number will
          be preserved.

                1   1134: contig of 1134 bp in length
             1135   1234: gap of unknown length
             1235   2288: contig of 1054 bp in length
             2289   2388: gap of unknown length
             2389   3489: contig of 1101 bp in length
             3490   3589: gap of unknown length
             3590   4603: contig of 1014 bp in length
             4604   4703: gap of unknown length
             4704   7152: contig of 2449 bp in length
             7153   7252: gap of unknown length
             7253   40260: contig of 33008 bp in length
            40261   40360: gap of unknown length
            40361   68687: contig of 28327 bp in length
            68688   68787: gap of unknown length
            68788   99752: contig of 30965 bp in length
            99753   99852: gap of unknown length
            99853   139204: contig of 39352 bp in length
           139205   139304: gap of unknown length
           139305   183920: contig of 44616 bp in length

FEATURES   Location/Qualifiers
     source   1..183920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone_lib="CalTech human BAC library D"
                     /clone="CTD-2159J19"

BASE COUNT   47125 a   45782 c   45629 g   44399 t   985 others
ORIGIN

Query Match   14.08;   Score 70.2;   DB 2;   Length 183920;
Best Local Similarity 46.2%;   Pred. No. 0.023;
Matches 216; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

Qy   18 GAGGCGGCGCAGCTGCGGAGCTCGGAGAAACAGGCGCGGCTTCCGCGCCCGGA 77
         ||| || ||||||| || ||| ||| ||| |||| | || |||||| ||||
Db 4205 GAGGGACNNCGGCGCGCGCCCCGGGGCGCGCCGCCCCCGCGCCCGCCACC 4146

Qy   78 CCCGGGCCCCAGATCATGATGCGTGCCCCGCCACCGCGCCCCCGGAGCGA 137
         || || ||| | |||||| ||| ||||||||||| || ||| |
Db 4145 GCCCCCGCGCCCCGCGCGCCCCGGGGTCGGGGGCGGGGGCCCCGCCGG 4086

Qy 138 TAGACGCGCCCCGGGGCCGGGGGTCTTGGAGTCCCCGGCCGCTGCCCCGGGAGGACC 197
         | ||| |||||||| || ||| | || | || |||| |||||||||||| ||
Db 4085 CGCCCCCGCGGCGCGCGCCACCCACCCGCCCCGGCCGGCCGCCCCCCG 4026

RESULT 8
AC084064/c
LOCUS   AC084064   197669 bp   DNA   linear   HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-467I17, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC084064
VERSION   AC084064.4 GI:13786424
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 197669)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 197669)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Apr 25, 2001 this sequence version replaced gi:11527453.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 0
            Center clone name: RPCI-11_467I17

            Summary Statistics
            Consensus quality: 158774 bases at least Q40
            Consensus quality: 174343 bases at least Q30
            Consensus quality: 180875 bases at least Q20
            Estimated insert size: 163300; agarose-fp estimation
            Estimated insert size: 195769; sum-of-contigs estimation
            Quality coverage: 5.93 in Q20 bases; agarose-fp estimation
            Quality coverage: 4.94 in Q20 bases; sum-of-contigs estimation
            NOTE: This is a 'working draft' sequence. It currently
            consists of 20 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.

                1   1297: contig of 1297 bp in length
             1298   1397: gap of unknown length
             1398   2963: contig of 1566 bp in length
             2964   3063: gap of unknown length

Qy 198 CACCCGCGCCTGCCCCCGATGCGTTGCAGTGGGGCGCCGCCATGGACGACGAGGATTACCCCGCAG 257
         | |||| || |||||| || || ||||| ||||||| || ||| ||||||||| |
Db 4025 CGCCCCCCCCGCCCCCCGCCGGCGCCGGGGCGGGGGCGCCGCCCCCCGCGCCCGCCCCCCGGGGGCGCCGCCCCCC 3966

Qy 258 CATGAACCCCCGCGCGGGCGAGCCTCCGTACAGCCGCGGCCCCTGCCAGAGCGCGCATG 317
         | | |||||| || || || | ||| || ||||| || ||| | |
Db 3965 GCCCCCCCCCGGCGGGCGCCGCGGCGNCCCCCGCGCCCNCCCCCGCGCC 3906

Qy 318 CTGCACTGCCCTACTGGAACACCTTCTGCGGCTGCCGCCATACCCTGCGTTCGTTCCAGCGAC 377
         | ||| ||| | || || | ||| || ||| ||||| |||| ||| |||||||
Db 3905 CCCCGCGCCCGCGCACCCGCGCCGCGCCCNCCCCCCCCCCCCGCGCCCGCGCC 3846

Qy 378 AGCCGCGCGGTTCATGAGGTCCGGCCGTCCTTCCTTCGGCGACCGGCGCCCTGCCCAGACACCAGC 437
         | ||||| | || | || |||| || | | | ||| || |||||||| |||||||| |
Db 3845 CCCCGCGCGCACCCGCGCCCCNCCCCCCCCCCCCGCGCCCGCGCCCTGCCCAGACACCAGC 3786

Qy 438 TATGCGCCGTGGGCCACCGCCCTTCCAGCCTTCGCCACCAAAAGACCTTGCGAC 485
         | || || |||| ||||||| ||||| ||| || || |||||||||||| |
Db 3785 CCCCCCCCCGNGCCCCCCCCCGNGCCNCCCNGGNCCNNCCCCCCGGC 3738

RESULT 9

```
Qy    462 AGCTTGCCACC 472
          ||| ||| ||
Db    612 CCCGNGCCCCC 602
```

```
LOCUS       AC091024               52804 bp    DNA     linear   HTG 24-MAR-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-845D20 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC091024
VERSION     AC091024.1  GI:13443184
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 52804)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-845D20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 52804)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,J.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lanazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/repeatMasker.html
            --------------
            Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------------
              Project Information
              Center project name: Li2650
              Center clone name: 845_D_20
            --------------
            * NOTE: This record contains 63 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            * ----------------
            *      1    704: contig of 704 bp in length
            *    705    804: gap of 100 bp
            *    805   1545: contig of 741 bp in length
            *   1546   1645: gap of 100 bp
```

FEATURES   Location/Qualifiers
     source   1..197669
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-467I17"
                     /clone_lib="RPCI human BAC library 11"

```
          3064   4476: contig of 1413 bp in length
          4477   4576: gap of unknown length
          4577   5741: contig of 1165 bp in length
          5742   5841: gap of unknown length
          5841   7080: contig of 1239 bp in length
          7081   7180: gap of unknown length
          7181   8247: contig of 1067 bp in length
          8247   8347: gap of unknown length
          8348  10242: contig of 1895 bp in length
         10243  10342: gap of unknown length
         10342  12553: contig of 2211 bp in length
         12554  12653: gap of unknown length
         12654  15376: contig of 2723 bp in length
         15377  15476: gap of unknown length
         15477  20581: contig of 5105 bp in length
         20582  20681: gap of unknown length
         20682  25105: contig of 4424 bp in length
         25106  25205: gap of unknown length
         25206  29870: contig of 4666 bp in length
         29871  29971: gap of unknown length
         29972  35022: contig of 5052 bp in length
         35023  35123: gap of unknown length
         35124  41897: contig of 6774 bp in length
         41898  41997: gap of unknown length
         41998  51710: contig of 9713 bp in length
         51711  51810: gap of unknown length
         51811  62531: contig of 10721 bp in length
         62532  62631: gap of unknown length
         62632  77530: contig of 14899 bp in length
         77531  77630: gap of unknown length
         77631 112421: contig of 34791 bp in length
        112422 112521: gap of unknown length
        112522 142859: contig of 30338 bp in length
        142860 142959: gap of unknown length
        142960 197669: contig of 54710 bp in length.
```

ORIGIN

BASE COUNT   53353 a   42514 c   43922 g   53642 t   4238 others

Query Match   13.9%; Score 69.6; DB 2; Length 197669;
Best Local Similarity   44.58%; Pred. No. 0.027;
Matches 165; Conservative   0; Mismatches 206; Indels   0; Gaps   0;

```
Qy   102 CCGGCCACCGGCGCCCACCACGGAGGAGCGAAAGCCCAGATAGACCCCCCGGCCCGGGGTC 161
              |||       ||| ||  | |||          |||||||||||||||||||||||||
Db   972 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGCGGGCCCCCGGGCCCGGGCCN 913

Qy   162 CTGGAGTCCCGCCCGGCGCCGCTGCCCGGCGGAGGACCCCCCACCCTGCCCCCCCGATGCTT 221
              ||| |||    ||||||||||||||||||||||||||||||  ||| | |||||||
Db   912 CGGCCNCCCCNCGNGNCCCCCCNCCCCCCCCCCCCCNNCNNNCGCCCNNGCCCNCCC 853

Qy   222 GCAGTGGGGCCCGCCCATGGACAGAGGGATTACCCGCAGCATGAACCCGCCGGGCCCAGC 281
              |||        ||| |||||||||||||||   ||||||||  |||||| |||||
Db   852 CCNCNGCCCCCCCGCNCCCNGCCCNCNGGNNCCCNCNGCCCCCCNCCCCCCCNC 793

Qy   282 CTCCTGTACAGACCCGCGCGCCCTGCAGAGCGCCATGCTGCACTGCCCCTTACTGGAACACC 341
              |||         | |||||||||||||||||||  ||| | ||||||| ||| ||
Db   792 CCCCCCNCCCCCCCCCCGGGNCCCCCCCCCCCCGNCCCNCNCCCCCCCCCC 733

Qy   342 TTCTCGCTGCCGCGCCCATACCCTGCCTTCTCAGGCGACACGCCCGGTTCATGAGCTCCGC 401
              ||||||||          | |||||||||||||||||||||||||| ||||| |
Db   732 CNNGNCCCGCGCGCCCCCGNGCCCCCCCNGCCCCCCCCNGCCCCCGCNCCCNCNGCNCC 673

Qy   402 TCCTTGCTTCCTGGCGAGCCAGCCTGCCCCAGAACACCAGCTATGCCCGGCCGCCACCGCCTCC 461
              |||  || ||||||||| ||| ||      |||||||| ||| |||| |||||| ||
Db   672 CCGCCNCCCCGGCCGGGNNCGCGCCCCNCCCCGGNCCCCCGCCCCCCCNCCC 613

Qy   462 AGCTTGCCACC 472
              ||| ||| ||
Db   612 CCCGNGCCCCC 602
```

```
*   1646   2386: contig of 741 bp in length
*   2387   2486: gap of 100 bp
*   2487   3205: contig of 719 bp in length
*   3206   3305: gap of 100 bp
*   3306   4062: contig of 757 bp in length
*   4063   4162: gap of 100 bp
*   4163   4875: contig of 713 bp in length
*   4876   4975: gap of 100 bp
*   4976   5728: contig of 753 bp in length
*   5729   5828: gap of 100 bp
*   5829   6574: contig of 746 bp in length
*   6575   6674: gap of 100 bp
*   6675   7408: contig of 734 bp in length
*   7409   7508: gap of 100 bp
*   7509   8237: contig of 729 bp in length
*   8238   8337: gap of 100 bp
*   8338   9066: contig of 729 bp in length
*   9067   9166: gap of 100 bp
*   9167   9912: contig of 746 bp in length
*   9913  10012: gap of 100 bp
*  10013  10731: contig of 719 bp in length
*  10732  10831: gap of 100 bp
*  10832  11587: contig of 756 bp in length
*  11588  11687: gap of 100 bp
*  11688  12424: contig of 737 bp in length
*  12425  12524: gap of 100 bp
*  12525  13276: contig of 752 bp in length
*  13277  13376: gap of 100 bp
*  13377  14078: contig of 702 bp in length
*  14079  14178: gap of 100 bp
*  14179  14922: contig of 744 bp in length
*  14923  15022: gap of 100 bp
*  15023  15766: contig of 744 bp in length
*  15767  15866: gap of 100 bp
*  15867  16600: contig of 734 bp in length
*  16601  16700: gap of 100 bp
*  16701  17456: contig of 756 bp in length
*  17457  17556: gap of 100 bp
*  17557  18318: contig of 762 bp in length
*  18319  18418: gap of 100 bp
*  18419  19152: contig of 734 bp in length
*  19153  19252: gap of 100 bp
*  19253  20022: contig of 770 bp in length
*  20023  20122: gap of 100 bp
*  20123  20824: contig of 702 bp in length
*  20825  20924: gap of 100 bp
*  20925  21652: contig of 728 bp in length
*  21653  21752: gap of 100 bp
*  21753  22475: contig of 723 bp in length
*  22476  22575: gap of 100 bp
*  22576  23333: contig of 758 bp in length
*  23334  23433: gap of 100 bp
*  23434  24188: contig of 755 bp in length
*  24189  24288: gap of 100 bp
*  24289  25022: contig of 734 bp in length
*  25023  25122: gap of 100 bp
*  25123  25878: contig of 756 bp in length
*  25879  25978: gap of 100 bp
*  25979  26710: contig of 732 bp in length
*  26711  26810: gap of 100 bp
*  26811  27543: contig of 733 bp in length
*  27544  27643: gap of 100 bp
*  27644  28363: contig of 720 bp in length
*  28364  28463: gap of 100 bp
*  28464  29211: contig of 748 bp in length
*  29212  29311: gap of 100 bp
*  29312  30074: contig of 763 bp in length
*  30075  30174: gap of 100 bp
*  30175  30934: contig of 760 bp in length
*  30935  31034: gap of 100 bp
*  31035  31767: contig of 733 bp in length
*  31768  31867: gap of 100 bp
*  31868  32632: contig of 765 bp in length

*  32633  32732: gap of 100 bp
*  32733  33462: contig of 730 bp in length
*  33463  33562: gap of 100 bp
*  33563  34304: contig of 742 bp in length
*  34305  34404: gap of 100 bp
*  34405  35156: contig of 752 bp in length
*  35157  35256: gap of 100 bp
*  35257  36004: contig of 748 bp in length
*  36005  36104: gap of 100 bp
*  36105  36845: contig of 741 bp in length
*  36846  36945: gap of 100 bp
*  36946  37671: contig of 726 bp in length
*  37672  37771: gap of 100 bp
*  37772  38526: contig of 755 bp in length
*  38527  38626: gap of 100 bp
*  38627  39341: contig of 715 bp in length
*  39342  39441: gap of 100 bp
*  39442  40191: contig of 750 bp in length
*  40192  40291: gap of 100 bp
*  40292  41046: contig of 755 bp in length
*  41047  41146: gap of 100 bp
*  41147  41869: contig of 723 bp in length
*  41870  41969: gap of 100 bp
*  41970  42689: contig of 720 bp in length
*  42690  42789: gap of 100 bp
*  42790  43552: contig of 763 bp in length
*  43553  43652: gap of 100 bp
*  43653  44405: contig of 753 bp in length
*  44406  44505: gap of 100 bp
*  44506  45257: contig of 752 bp in length
*  45258  45357: gap of 100 bp
*  45358  46090: contig of 733 bp in length
*  46091  46190: gap of 100 bp
*  46191  46919: contig of 729 bp in length
*  46920  47019: gap of 100 bp
*  47020  47743: contig of 724 bp in length
*  47744  47843: gap of 100 bp
*  47844  48616: contig of 773 bp in length
*  48617  48716: gap of 100 bp
*  48717  49438: contig of 722 bp in length
*  49439  49538: gap of 100 bp
*  49539  50290: contig of 752 bp in length
*  50291  50390: gap of 100 bp
*  50391  51126: contig of 736 bp in length
*  51127  51226: gap of 100 bp
*  51227  51966: contig of 740 bp in length
*  51967  52066: gap of 100 bp
*  52067  52804: contig of 738 bp in length.
```

FEATURES   Location/Qualifiers
     source   1..52804
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-845D20"
                     /clone_lib="RPCI-11 Human Male BAC"

BASE COUNT   11285 a   12265 c   9180 g   12489 t   7585 others
ORIGIN

Query Match   13.9%;   Score 69.4;   DB 2;   Length 52804;
Best Local Similarity   42.0%;   Pred. No. 0.04;
Matches 193;   Conservative 0;   Mismatches 267;   Indels 0;   Gaps 0;

```
Qy     13 GAAGTGAGCGCGGCCGGCGCAGCTGCGAGCGTTCGGAGAAAACAGGCGCCGCGGGCTCCGGCGCCGG 72
          ||||  ||||| ||||| |||| || |  | |||| ||| ||     || | || ||| ||| ||| ||
Db  27056 GAAGGGGCGGCGGGAGCGGGGCTGNGGNGCCNGNCCTCNCNCNTTCTTCNTCTGCCCCCCCC 27115

Qy     73 CCGGACCCGCGGCCCCGAGATCATGATGACTTGCCGCCACCGCCGCCACCGACGGAGGAAGC 132
          |||| ||| ||| ||| || |  |  |  | |  |  ||  ||  |    |  || ||||
Db  27116 CCGGNCCCGCGCGGCGCNCCGTCCCCGCCCTCCCNCNCNCNNNNNCCCCC 27175

Qy    133 CCAGATAGACGCCCGGCGCCGGCGGCCCCGGGCCCGGCGCCGGCGCCGGGGTCCTGGCGCTGCGGGCCGCGGGCGAG 192
          ||    ||||  ||||  ||||   ||||   ||||  |||| | |  ||||       ||  ||  ||   ||  ||
```

```
Db  27176  CCTCCNNNNCNCCCCCCCCCCCCCCCNCCCCCCGCCCCCCCCCCNNCCCCNCCCCGCCCC  27235
                 |  | | ||||
Qy    193  GACCCGCACCCTGCCCGCCCGGATCGCTTGCAGTGTGGGGCCCGGCCCATGGACAGGGATTACC  252
Db  27236  NNCNCCCACCCCCCNCCCCCCCCCCCCCCCCCCCCNNCCCCCNNCCCCCCCCCCCCCC  27295
                 || || ||||
Qy    253  CGCAGGCATGAACCCCCCCCCGCGGGGGGGGGAGCCCTCCTGTACAGCCCGCCCCTGCCAGAGCG  312
                 ||||  ||| ||
Db  27296  CCNNNCNCNCCCCCCCCCNCCCNCNNCCCCCCCCCCCNCCCCCCCCCCCCCNNC  27355
                 |
Qy    313  CCATGCTGCACTGCCCTACTGGAACACCTTCTCGCGTCGCTGCGCCGCCATACCCTGCCTTCTCCA  372
                 || | || ||||
Db  27356  CCCCCCCCCCCCCCNNNNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  27415
                 || |||
Qy    373  GCGACAGCCGCCGCCGGTTCATGAGCTCCGCCTCCTTCCTCGGCACCCAGCCCTGCCCCAGACA  432
                 |||  ||||
Db  27416  CCCCCNNCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  27475
                 |||||||  || |||
Qy    433  CCAGCTATGCCCGTGGCCACCGCCTCCAGCTTGCCACC  472
                 ||  || |||
Db  27476  CCCNCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC  27515


RESULT 10
ACC079424/c

LOCUS       AC079424         209887 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-111E8, WORKING DRAFT
            SEQUENCE, 47 unordered pieces.
ACCESSION   AC079424
VERSION     AC079424.1  GI:9958036
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209887)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Mouse
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 209887)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_111E8
            -----
            Summary Statistics
            Consensus quality: 116537 bases at least Q40
            Consensus quality: 149032 bases at least Q30
            Consensus quality: 164980 bases at least Q20
            Estimated insert size: 153300; agarose-fp estimation
            Estimated insert size: 205287; sum-of-contigs estimation
            Quality coverage: 3.52 in Q20 bases; agarose-fp estimation
            Quality coverage: 2.63 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 47 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *       1        1072: contig of 1072 bp in length
            *    1073        1172: gap of unknown length
            *    1173        2240: contig of 1068 bp in length
            *    2241        2340: gap of unknown length
```

```
                  2341:   3424: contig of 1084 bp in length
                  3425:   3524: gap of unknown length
                  3525:   4592: contig of 1068 bp in length
                  4593:   4692: gap of unknown length
                  4693:   5715: contig of 1023 bp in length
                  5716:   5815: gap of unknown length
                  5816:   6918: contig of 1103 bp in length
                  6919:   7018: gap of unknown length
                  7019:   8127: contig of 1109 bp in length
                  8128:   8227: gap of unknown length
                  8228:   9282: contig of 1055 bp in length
                  9283:   9382: gap of unknown length
                  9383:  10474: contig of 1092 bp in length
                 10475:  10574: gap of unknown length
                 10575:  11692: contig of 1118 bp in length
                 11693:  11792: gap of unknown length
                 11793:  12899: contig of 1107 bp in length
                 12900:  12999: gap of unknown length
                 13000:  14001: contig of 1002 bp in length
                 14002:  14101: gap of unknown length
                 14102:  15119: contig of 1018 bp in length
                 15120:  15219: gap of unknown length
                 15220:  16419: contig of 1200 bp in length
                 16420:  16519: gap of unknown length
                 16520:  18558: contig of 2039 bp in length
                 18559:  18658: gap of unknown length
                 18659:  20567: contig of 1909 bp in length
                 20568:  20667: gap of unknown length
                 20668:  23000: contig of 2333 bp in length
                 23001:  23100: gap of unknown length
                 23101:  24648: contig of 1548 bp in length
                 24649:  24748: gap of unknown length
                 24749:  26966: contig of 2218 bp in length
                 26967:  27066: gap of unknown length
                 27067:  28799: contig of 1733 bp in length
                 28800:  28899: gap of unknown length
                 28900:  31223: contig of 2324 bp in length
                 31224:  31323: gap of unknown length
                 31324:  33361: contig of 2038 bp in length
                 33362:  33461: gap of unknown length
                 33462:  35805: contig of 2344 bp in length
                 35806:  35905: gap of unknown length
                 35906:  38759: contig of 2854 bp in length
                 38760:  38859: gap of unknown length
                 38860:  40679: contig of 1820 bp in length
                 40680:  40779: gap of unknown length
                 40780:  43413: contig of 2634 bp in length
                 43414:  43513: gap of unknown length
                 43514:  46086: contig of 2573 bp in length
                 46087:  46186: gap of unknown length
                 46187:  49745: contig of 3559 bp in length
                 49746:  49845: gap of unknown length
                 49846:  53931: contig of 4086 bp in length
                 53932:  54031: gap of unknown length
                 54032:  55489: contig of 1458 bp in length
                 55490:  55589: gap of unknown length
                 55590:  60845: contig of 5256 bp in length
                 60846:  60945: gap of unknown length
                 60946:  64573: contig of 3628 bp in length
                 64574:  64673: gap of unknown length
                 64674:  70953: contig of 6280 bp in length
                 70954:  71053: gap of unknown length
                 71054:  76194: contig of 5041 bp in length
                 76195:  76294: gap of unknown length
                 76295:  81178: contig of 4984 bp in length
                 81179:  81278: gap of unknown length
                 81279:  86878: contig of 5600 bp in length
                 86879:  86978: gap of unknown length
                 86979:  88585: contig of 1607 bp in length
                 88586:  88685: gap of unknown length
                 88686:  97517: contig of 8832 bp in length
                 97518:  97617: gap of unknown length
                 97618: 104566: contig of 6949 bp in length
```

```
RESULT 11
AC084799/c
LOCUS       AC084799    303091 bp    DNA    linear    HTG 17-NOV-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-197M9, WORKING DRAFT
            SEQUENCE, 101 unordered pieces.
ACCESSION   AC084799
VERSION     AC084799.1  GI:11192127
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
```

```
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 303091)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 303091)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     ------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            ------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_197M9
            ------
            Summary Statistics
            Consensus quality:    152568 bases at least Q40
            Consensus quality:    175579 bases at least Q30
            Consensus quality:    186949 bases at least Q20
            Estimated insert size: 198300; agarose-fp estimation
            Estimated insert size: 293091; sum-of-contigs estimation
            Quality coverage:  4.85 in Q20 bases; agarose-fp estimation
            Quality coverage:  3.28 in Q20 bases; sum-of-contigs estimation
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 101 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
```

```
FEATURES             Location/Qualifiers
     source          1..209887
                     /organism="Mus musculus"
                     /clone="RP23-111E8"
                     /clone_lib="RPCI mouse BAC library 23"
                     /chromosome="16"
                     /db_xref="taxon:10090"
```

```
BASE COUNT    52682 a   45440 c   51398 g   52137 t   8230 others
ORIGIN
```

```
*       1     1116: contig of 1116 bp in length
*    1117     1216: gap of unknown length
*    1216     1320: contig of 1320 bp in length
*    1217     1320: gap of unknown length
*    2536     2636: contig of 1421 bp in length
*    2537     2636: gap of unknown length
*    2637     4057: contig of 1421 bp in length
*    4058     4157: gap of unknown length
*    4158     5245: contig of 1088 bp in length
*    5246     5345: gap of unknown length
*    5346     6461: contig of 1116 bp in length
*    5462     6461: gap of unknown length
*    6462     6561: contig of 1116 bp in length
*    6562     7725: gap of unknown length
*    7726     7825: contig of 1164 bp in length
*    7826     9143: gap of unknown length
*    9144     9243: contig of 1318 bp in length
*    9244    10492: gap of unknown length
*   10493    10592: contig of 1249 bp in length
*   10593    11666: gap of unknown length
*   11667    11766: contig of 1074 bp in length
*   11767    12950: gap of unknown length
*   12951    13050: contig of 1184 bp in length
*   13051    14285: gap of unknown length
*   14286    14385: contig of 1235 bp in length
*   14386    15478: gap of unknown length
*   15479    15578: contig of 1093 bp in length
*   15579    16779: gap of unknown length
*   16780    16879: contig of 1201 bp in length
*   16880    18278: gap of unknown length
*   18279    18378: contig of 1399 bp in length
*   18379    19452: gap of unknown length
*   19453    19552: contig of 1074 bp in length
*   19553    21012: gap of unknown length
*   21013    21112: contig of 1460 bp in length
*   21113    22169: gap of unknown length
*   22170    22269: contig of 1057 bp in length
*   22270    23367: gap of unknown length
*   23368    23467: contig of 1098 bp in length
*   23468    24559: gap of unknown length
*   24559          : contig of 1092 bp in length
```

```
Qy     48  CAGGCGCCGCGGCGGTCCGCGCCGCCGGAGGATCATGATGCTGCCGCCA  107
               |||  ||| || | ||  ||||
Db   6710  CCNNCCCCNNNNCCCCCNNNNCCNNCNNNCNNCCCCNCCCCCCCCNNCCCCC  6651

Qy    108  CCGGCGGCCACCAGGAGCGGAGAAGCCCAGATAGAGACCCCGCCGGGTCCTGGAG  167
               |||| |  |||
Db   6650  CCCCCCCCCCCNNCNNCNNNCCNCCCCCCCCCNCNCCCCCCCNCCCCNN  6591

Qy    168  TCCCGCCGCCTTGCTGCCCGGCCGAGGACCCCACCCGCCGCCTGCCCGCCGATGCTTGCAGTG  227
               || |||| |||  ||||| |
Db   6590  NCCCCCCCCCCCCNNNNCCNCCCCCCGCCCCCCCCCCCCCNNCCCCCCNNC  6531

Qy    228  GGGCCCCGCCCCATGAGACAGGGGATTACCCGCAGCATGAAACCCCGCGGGCCAGCCTCCTG  287
               ||| ||  |||
Db   6530  CCCCCCNCCCCCNCCCNNNCCCCCCCCCCNNNNCCCCCCCCCCCCCCCCN  6471

Qy    288  TACAGACCCGCGCCGCCGCGCCAGAGACGGGCCCATGCTGCACTGCCCCTACTGGGAACACCTTCTG  347
               || | |||  |||
Db   6470  CCCNCNCCCCCCCCCCCNNCCCCNCCCCCCCCCCCNNCCCCCCNN  6411

Qy    348  CTGCCGGCCATACCCTGCCTTCTCCAGCGACAGACCGCGCGGTTCATGAGCTCCGCCCTCCTTC  407
               |||  |||  |||||||
Db   6410  CCNCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCNNNCN  6351

Qy    408  CTCGGCAGCCAGCCCTGCCCAGACACCAGCTATGCCCCGTGGCCACCCGCCTCCAGCTTG  467
               |||| ||  |||  ||||  |||||
Db   6350  CCCCCCNCCCCCCCCCCCCCCCCCNNNCCCCCCCCCCNNNCCCCC  6291

Qy    468  CCACC  472
           |||||
Db   6290  CCCCC  6286
```

```
Query Match        13.8%;  Score 69.2;  DB 2;  Length 209887;
Best Local Similarity 40.0%;  Pred. No. 0.031;  DB 2;
Matches 170;  Conservative 0;  Mismatches 255;  Indels 0;  Gaps 0;
```

```
*  104567   104666: gap of unknown length
*  104667   111806: contig of 7140 bp in length
*  111807   111906: gap of unknown length
*  111907   120059: contig of 8153 bp in length
*  120060   120159: gap of unknown length
*  120160   131317: contig of 11158 bp in length
*  131318   131417: gap of unknown length
*  131418   146920: contig of 15503 bp in length
*  146921   147020: gap of unknown length
*  147021   160830: contig of 13810 bp in length
*  160831   160930: gap of unknown length
*  160931   181738: contig of 20808 bp in length
*  181739   181838: gap of unknown length
*  181839   197846: contig of 16008 bp in length
*  197847   197946: gap of unknown length
*  197947   209887: contig of 11941 bp in length.
```

```
*   24560   24659: gap of unknown length
*   24660   25825: contig of 1166 bp in length
*   25826   25925: gap of unknown length
*   25926   27036: contig of 1111 bp in length
*   27037   27136: gap of unknown length
*   27137   28245: contig of 1109 bp in length
*   28246   28345: gap of unknown length
*   28346   29656: contig of 1311 bp in length
*   29657   29756: gap of unknown length
*   29757   30883: contig of 1127 bp in length
*   30884   30983: gap of unknown length
*   30984   32125: contig of 1142 bp in length
*   32126   32225: gap of unknown length
*   32226   33397: contig of 1172 bp in length
*   33398   33497: gap of unknown length
*   33498   34645: contig of 1148 bp in length
*   34646   34745: gap of unknown length
*   34746   35894: contig of 1149 bp in length
*   35895   35994: gap of unknown length
*   35995   37156: contig of 1162 bp in length
*   37157   37256: gap of unknown length
*   37257   38396: contig of 1140 bp in length
*   38397   38496: gap of unknown length
*   38497   39570: contig of 1074 bp in length
*   39571   39670: gap of unknown length
*   39671   40852: contig of 1182 bp in length
*   40853   40952: gap of unknown length
*   40953   42116: contig of 1164 bp in length
*   42117   42216: gap of unknown length
*   42217   43372: contig of 1156 bp in length
*   43373   43472: gap of unknown length
*   43473   44603: contig of 1131 bp in length
*   44604   44703: gap of unknown length
*   44704   45924: contig of 1221 bp in length
*   45925   46024: gap of unknown length
*   46025   47174: contig of 1150 bp in length
*   47175   47274: gap of unknown length
*   47275   48399: contig of 1125 bp in length
*   48400   48499: gap of unknown length
*   48500   49969: contig of 1470 bp in length
*   49970   50069: gap of unknown length
*   50070   51135: contig of 1066 bp in length
*   51136   51235: gap of unknown length
*   51236   52690: contig of 1455 bp in length
*   52691   52790: gap of unknown length
*   52791   53920: contig of 1130 bp in length
*   53921   54020: gap of unknown length
*   54021   55101: contig of 1081 bp in length
*   55102   55201: gap of unknown length
*   55202   56324: contig of 1123 bp in length
*   56325   56424: gap of unknown length
*   56425   57583: contig of 1159 bp in length
*   57584   57683: gap of unknown length
*   57684   58822: contig of 1139 bp in length
*   58823   58922: gap of unknown length
*   58923   60547: contig of 1625 bp in length
*   60548   60647: gap of unknown length
*   60648   61743: contig of 1096 bp in length
*   61744   61843: gap of unknown length
*   61844   62998: contig of 1155 bp in length
*   62999   63098: gap of unknown length
*   63099   64220: contig of 1122 bp in length
*   64221   64320: gap of unknown length
*   64321   65398: contig of 1078 bp in length
*   65399   65498: gap of unknown length
*   65499   66762: contig of 1264 bp in length
*   66763   66862: gap of unknown length
*   66863   68025: contig of 1163 bp in length
*   68026   68125: gap of unknown length
*   68126   69400: contig of 1275 bp in length
*   69401   69500: gap of unknown length
*   69501   70718: contig of 1218 bp in length
*   70719   70818: gap of unknown length
```

```
Query Match           Score 69.2;  DB 2;   Length 303091;
Best Local Similarity 13.8%;   Pred. No. 0.029;
Matches 179;  Conservative 0;  Mismatches 270;  Indels 0;  Gaps 0;

Qy     52 CGCCGGGGGGCGTTCCCGGCGCGGACCCGGGCCGGGACCCGGAGATCAATGATGCTGCCCACCGC 111
          ||| ||| |||  || |
Db  57095 CCCCCCCCNNCCCCCCCCCCNNCCCCCCCNNCCCCCCCNN 57036

Qy    112 CGCCACCACGGAGCGAGCGAGAAGCCCAGATAGACGCCCCGGGGGTCCTGGAGTCCC 171
          || |||| ||
Db  57035 CCCCNCCCCNNCCCCCCCCCNNNNCCCCCCCCCCC 56976

Qy    172 GCCGGCTGCTTGCCCGGCGCCGAGGACCCCACCCCGGGGGCCGGCGATGCTGCACGTGGGGC 231
          |||| |||
Db  56975 CCCCCNCCCCCCCCCCCNNCCCCCNNCCCCCCCCCCCCCCC 56916

Qy    232 CCGCCCATGGACAGGGATTACCCGCCAGCATGAACCCCCGCCCAGCCGCCTCCCCGTACA 291
          ||| | |||
Db  56915 NNCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCCCCCCC 56856

Qy    292 GCCCGCGCGCCCCTGCAGAGCGGCCATGCTGCACTGCCCCTACTGGAAACACCTTCTCGGCTGC 351
          ||| || || |
Db  56855 CCCCCCCCCCCCCCCCCCNNCCCCCCNNCCCCCCCCCCC 56796

Qy    352 CGCCATACCCTGCCTTCTTCCAGCGACGAGCCGCGCCGTTCATGAGCTCCCCTTCTTCCTCG 411
          ||| ||| || | | |||
Db  56855 CCCCCCCCCCNNCCNNCCCCCCCCCCNNCCCCCCCCCCC 56796
```

```
Db   56795  CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCNNCCNNNNNNCCCCNCCCCNCC 56736
              ||| ||||||||||||||||||||||||| |||||| ||  ||| ||| ||
Qy     412  GCAGCCAGCCCGTGCCCAGACACCAGCTATGCCCCGTGGCCACCGCCTCCAGCTTGCCAC 471

Db   56735  CCCCCNNNCCCNNCCCCNNNNCNCCNCCNCCNNNCCCCCCNNNCCNNCCCCNNCCNCCCCC
              |||| ||| || |||| |||| | || ||| || ||| ||  || ||||| || |||||
Qy     472  CAAAGACCTGCGACTTTGCTCAGGACTCC 500

Db   56675  CCCCCCCCCCCCCCCCCCCCCCCNNCCCNNNNNNCCCCCNNC 56647
```

RESULT 12

LOCUS   AC096161   66507 bp   DNA   linear   HTG 20-DEC-2001
DEFINITION   Rattus norvegicus clone CH230-11D21, *** SEQUENCING IN PROGRESS
             ***, 44 unordered pieces.
ACCESSION   AC096161
VERSION   AC096161.2 GI:17943847
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 66507)
AUTHORS   Mizny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
             Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
             Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
             Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
             Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
             Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
             Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
             Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
             Joudah,S., Karlson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
             Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
             Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
             Morgan,M., Morris,S., Moser,M., Neal,J., Newton,J., Newtson,N.,
             Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
             Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,N., Pace,A., Payton,B.,
             Peery,J., Perez,L., Peters,L., Pickens,I., Primus,E., Pu,L.L.,
             Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
             Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
             Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
             Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
             Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
             Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
             Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
             Watlington,S., Williams,S., Williamson,A., Wleczyk,R., Wooden,S.,
             Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 66507)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Dec 20, 2001 this sequence version replaced gi:15627781.

             --------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
             ---------- Project Information
               Center project name: GEOC
               Center clone name: CH230-11D21
             --------------- Summary Statistics
             Assembly program: Phrap; version 0.990329First call to
             findPhraplist
               Consensus quality: 39129 bases at least Q40
               Consensus quality: 47892 bases at least Q30
               Consensus quality: 54035 bases at least Q20
               Estimated insert size: 23598; sum-of-contigs estimation
               Quality coverage: 0x in Q20 bases; agarose-fp estimation
               Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation
             ----
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 44 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *
             *   1 2381:   contig of 2381 bp in length
             * 2382 2481:   gap of unknown length
             * 2482 3972:   contig of 1491 bp in length
             * 3973 4072:   gap of unknown length
             * 4073 5565:   contig of 1493 bp in length
             * 5566 5665:   gap of unknown length
             * 5666 7271:   contig of 1606 bp in length
             * 7272 7371:   gap of unknown length
             * 7372 8877:   contig of 1506 bp in length
             * 8878 8977:   gap of unknown length
             * 8978 11392:   contig of 2415 bp in length
             * 11393 11492:   gap of unknown length
             * 11493 13131:   contig of 1639 bp in length
             * 13132 13231:   gap of unknown length
             * 13232 14331:   contig of 1100 bp in length
             * 14332 14431:   gap of unknown length
             * 14432 16776:   contig of 2345 bp in length
             * 16777 16876:   gap of unknown length
             * 16877 18398:   contig of 1522 bp in length
             * 18399 18498:   gap of unknown length
             * 18499 20030:   contig of 1532 bp in length
             * 20031 20130:   gap of unknown length
             * 20131 22510:   contig of 2380 bp in length
             * 22511 22610:   gap of unknown length
             * 22611 23705:   contig of 1095 bp in length
             * 23706 23805:   gap of unknown length
             * 23806 25008:   contig of 1203 bp in length
             * 25009 25108:   gap of unknown length
             * 25109 26645:   contig of 1537 bp in length
             * 26646 26745:   gap of unknown length
             * 26746 28658:   contig of 1913 bp in length
             * 28659 28758:   gap of unknown length
             * 28759 30510:   contig of 1752 bp in length
             * 30511 30610:   gap of unknown length
             * 30611 32461:   contig of 1851 bp in length
             * 32462 32561:   gap of unknown length
             * 32562 34076:   contig of 1515 bp in length
             * 34077 34176:   gap of unknown length
             * 34177 35359:   contig of 1183 bp in length
             * 35360 35459:   gap of unknown length
             * 35460 36541:   contig of 1082 bp in length
             * 36542 36641:   gap of unknown length
             * 36642 37693:   contig of 1052 bp in length
             * 37694 37793:   gap of unknown length
             * 37794 39307:   contig of 1514 bp in length
```

```
BASE COUNT    16704 a    15295 c    12208 g    17205 t     5095 others
ORIGIN

FEATURES             Location/Qualifiers
     source          1..66507
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-11D21"
```

```
Query Match       13.8%;  Score 69;  DB 2;  Length 66507;
Best Local Similarity   47.7%;  Pred. No. 0.043;
Matches  201;  Conservative   0;  Mismatches  220;  Indels   0;  Gaps    0;

Qy         52 CGCCGCGGGGGTTCCGCCCCGCGGGCCCCGAGATCATGATGTGTGCCCGCCACCC 111
              ||||  |  |||||| ||  ||||||||||||||| ||| | || ||| ||| |||
Db      35649 CGCCCCCACCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35708

Qy        112 CGCCACCACGGAGGCGAGAGCCCAGATAGACCCCCCCCGGGGTCCTGGAGTCC 171
              |||| ||| ||| |||                  | |||
Db      35709 CCCCCCCCCACCGCCCCCCCCG CCCCCCCCCCCCCCCCCCCCCG 35768

Qy        172 GCCGGCCTGCCTGCCGCGGGCCGAGGACACCCACCCGCCGATGCTTGCAGTGGGGGC 231
              || | |  |||| |||  |||| |||||||||||||||| |  |||| | |
Db      35769 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCC 35828

Qy        232 CCGCCATGGGACAGGGATTACCCGCAGGATGAACCCCCGCGGGGGCAGCCTTCTGTACA 291
              || | ||||||                |||| || ||| ||
Db      35829 CCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35888

Qy        292 GCCCGGCGCGCCCCTGCAGAGCGCCCATGCTGCACTGCAGGCCCTTACTGAAACACCTTCTGCTGC 351
              ||||||||||||||||||||||||||||  |||   |  ||| | || | ||||
Db      35889 CCCCCCCGCGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 35948
```

```
RESULT 13
AC034105/c

LOCUS       AC034105              158167 bp    DNA     linear    HTG 01-AUG-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-170L3, WORKING DRAFT
            SEQUENCE, 5 ordered pieces.
ACCESSION   AC034105
VERSION     AC034105.5  GI:15055291
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158167)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158167)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 1, 2001 this sequence version replaced gi:13786407.
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible as part
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *
            *           1     5107: contig of 5107 bp in length
            *        5108     5207: gap of unknown length
            *        5208    17738: contig of 12531 bp in length
            *       17739    17838: gap of unknown length
            *       17839    22211: contig of 4373 bp in length
            *       22212    22311: gap of unknown length
            *       22312    31009: contig of 8698 bp in length
            *       31010    31109: gap of unknown length
```

```
Qy        352 CGGCCATACCCTTGCCTTTCTCCAGCGACGACAGCCCGCCCGGTTCATGAGCTTCGGCTCCTTCTCG 411
              |||  | |  |||| ||  || ||||||||||||  ||||| ||||  |  |  || ||| | |||
Db      35949 CCCCCCCCCCCCCCCCC CCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCGACCCCGTCG 36008

Qy        412 GCAGCCAGCCCTGCCCAGACACCAGCTATGCCCCCCGTGGGCCCACCGCCGCCTTCCAGCTTGCCAC 471
              ||     ||| | |  |  || || || |||||    ||||||  ||||   |  |||| |  ||||
Db      36009 CCGCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 36068

Qy        472 C 472
              |
Db      36069 C 36069
```

```
*       39308   39407: gap of unknown length
*       39408   40733: contig of 1326 bp in length
*       40734   40833: gap of unknown length
*       40834   41882: contig of 1049 bp in length
*       41883   41982: gap of unknown length
*       41983   43061: contig of 1079 bp in length
*       43062   43161: gap of unknown length
*       43162   44564: contig of 1403 bp in length
*       44565   44664: gap of unknown length
*       44665   45743: contig of 1079 bp in length
*       45744   45843: gap of unknown length
*       45844   46943: contig of 1100 bp in length
*       46944   47043: gap of unknown length
*       47044   48095: contig of 1052 bp in length
*       48096   48195: gap of unknown length
*       48196   49771: contig of 1576 bp in length
*       49772   49871: gap of unknown length
*       49872   50992: contig of 1121 bp in length
*       50993   51092: gap of unknown length
*       51093   52404: contig of 1312 bp in length
*       52405   52504: gap of unknown length
*       52505   53633: contig of 1129 bp in length
*       53634   53733: gap of unknown length
*       53734   54776: contig of 1043 bp in length
*       54777   54876: gap of unknown length
*       54877   56230: contig of 1354 bp in length
*       56231   56330: gap of unknown length
*       56331   57844: contig of 1514 bp in length
*       57845   57944: gap of unknown length
*       57945   58982: contig of 1038 bp in length
*       58983   59082: gap of unknown length
*       59083   60411: contig of 1329 bp in length
*       60412   60511: gap of unknown length
*       60512   61671: contig of 1160 bp in length
*       61672   61771: gap of unknown length
*       61772   62805: contig of 1034 bp in length
*       62806   62905: gap of unknown length
*       62906   64267: contig of 1362 bp in length
*       64268   64367: gap of unknown length
*       64368   65404: contig of 1037 bp in length
*       65405   65504: gap of unknown length
*       65505   66507: contig of 1003 bp in length.
```

RESULT 14

```
LOCUS       CNS01BNS                840 bp    mRNA      linear     PLN 02-SEP-1999
DEFINITION  Botrytis cinerea strain T4 cDNA library under conditions of
            nitrogen deprivation.
ACCESSION   AL114464
VERSION     AL114464.1  GI:5829083
KEYWORDS    cDNA library; nitrogen deprivation.
SOURCE      Botryotinia fuckeliana.
ORGANISM    Botryotinia fuckeliana
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE   1 (bases 1 to 840)
AUTHORS     Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE       Direct Submission
JOURNAL     Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
            78026 Versailles, France
REFERENCE   2 (bases 1 to 840)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
            CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The cDNA library to be analyzed within the framework of this
            project was created using a Botrytis cinerea strain which was grown
            under conditions of nitrogen deprivation, which is the normal
            situation for B. cinerea during its development on its host plant.
            The library was produced in an oriented direction, in the pBSII
            vector.
FEATURES             Location/Qualifiers
     source          1..840
                     /organism="Botryotinia fuckeliana"
                     /strain="T4"
                     /db_xref="taxon:40559"
                     /note="Genoscope sequence ID : W45A0031"
BASE COUNT     24 a    700 c    75 g    33 t    8 others
ORIGIN
```

Query Match   13.6%; Score 68; DB 8; Length 840;
Best Local Similarity 47.5%;
Matches 200; Conservative 0; Pred. No. 0.17; Mismatches 221; Indels 0; Gaps 0;

```
Qy     52 CGCCGCGGGGGCTGCCGACGGCCGGAGCCCGGGGGCCCGGGAGATCACGAGCTGCCGCCACGC 111
          |||  || || |  || || || ||   |   ||||| | |||   |  ||| |   | |||
Db    387 CCCCCCGGGGGGCCGGGCCGGCCCGCCGGGGCGGGGGCCCCGGGGGGTCCTGGAGTCC 446

Qy    112 CGCCACCACGGAGCGGAGGAAGCCCAGATAGACGCGCCCGGGGGTCCTGGAGTCC 171
          |||   || |||  ||   || |  ||  |  ||  |||  | |  |||
Db    447 CCCCCCGGCCCCGCCCCCCCCCCCCCCCGCCCCGGGGGCCCCGGGGCCCCCCC 506

Qy    172 GCCGGCTGCTGCCGCGGCGGAGGACCCCACCCCGGCCCCATGCTTTGCAGTGGGGC 231
          ||  || | | |   || |  |||  | ||   || ||   |  |   |  ||||
Db    507 CCCCCCGGCCCCCCCGGAGGACCCCACCCCCGCCCCCCCCCCGGATGCTTGCAGTGCTGGGGC 566

Qy    232 CCGGCCATGGACAGGGATTACCCGCACCATGAACCCCGCCGCCGGGGCAGCCTCCTGTACA 291
          |||||||||||||||||||||||||  |  |   |||||   | || | ||| || || ||
Db    567 CCGGCCATGGACAGGGATTACCCGCAGCATGAACCCCCCCCCCGGGGGGGCAGCCTCCTGTACA 626

Qy    292 GCCGCCCGTTCTTCCAGGCACGACGCGCCGGTTCATGAGCTCCCGCCTCCTTCCTCG 351
          | |    |  |   |   ||| ||| || || |||||| |  ||| | |   |   ||
Db    627 CCCCCCCCCTGCAGAGCGCCATGCTGCACTGCCCTACTGGAACACCTTCTCGGCTGC 686

Qy    352 CGCCATACCCTGCCTTCTCCAGCGACGAGCGCCGGTTCATGAGCTCCCGCCTCCTTCCTCG 411
          ||| | | || |  |||||| ||  | ||   || || ||   |  ||| | |  |  | ||
Db    687 CGCCCCCGCCCCCCTCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCGGCCCCCCC 746

Qy    412 GCAGCCAGCCCTGCCCAGACACCAGCTATGCCCCCGGTGGCCACCGCTTCCAGCTTGCCAC 471
          ||| ||| || ||||||| |||| ||||||||    || |   || || ||||  |||||
Db    747 GCAGCCAGCCCCGCCCCGCCCCCCCCCCCCCCCCCCCGGCCCCGGCTCCCAGCCNCCCCCC 806

Qy    472 C 472
          |
Db    807 C 807
```

FEATURES
```
     *  31110   158167: contig of 127058 bp in length.
                        Location/Qualifiers
        1..158167
     source
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-170L3"
BASE COUNT   48614 a   32413 c   31779 g   44829 t   532 others
ORIGIN
```

Query Match   13.8%; Score 69; DB 2; Length 158167;
Best Local Similarity 47.7%;
Matches 201; Conservative 0; Pred. No. 0.036; Mismatches 220; Indels 0; Gaps 0;

RESULT 15

```
LOCUS       AC105671           143435 bp    DNA       linear     HTG 09-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-52J8, *** SEQUENCING IN PROGRESS ***,
            73 unordered pieces.
ACCESSION   AC105671
VERSION     AC105671.1  GI:18092894
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 143435)
AUTHORS     Muzny,D.M., Adams,C., Amaratunge,H.C., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Deshazo,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
```

FEATURES
```
                        Location/Qualifiers
        1..840
     source
                     /organism="Botryotinia fuckeliana"
                     /strain="T4"
                     /db_xref="taxon:40559"
                     /note="Genoscope sequence ID : W45A0031"
```

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Harris,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Soderorgren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

REFERENCE   2 (bases 1 to 143435)
AUTHORS   Unpublished
JOURNAL   Direct Submission
TITLE   Worley,K.C.
JOURNAL   Direct Submission
TITLE   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA

COMMENT   ----------- Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu/
           Contact: hgsc-help@bcm.tmc.edu
           ----------- Project Information
           Center project name: GNMH
           Center clone name: CH230-52J8
           ----------- Summary Statistics
           Assembly program: Phrap; version 0.990329First call to
           findPhrapList
           Consensus quality: 107670 bases at least Q40
           Consensus quality: 117102 bases at least Q30
           Consensus quality: 123580 bases at least Q20
           Estimated insert size: 98010; sum-of-contigs estimation
           Quality coverage: 0x in Q20 bases; agarose-fp estimation
           Quality coverage: 1.3x in Q20 bases; sum-of-contigs estimation
           ----
       * NOTE: Estimated insert size may differ from sequence length
       * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
       * NOTE: This is a 'working draft' sequence. It currently
       * consists of 73 contigs.The true order of the pieces
       * is not known and their order in this sequence record is
       * arbitrary. Gaps between the contigs are represented as
       * runs of N, but the exact sizes of the gaps are unknown.
       * This record will be updated with the finished sequence
       * as soon as it is available and the accession number will
       * be preserved.

*	1	4845:	contig of 4845 bp in length
*	4846	4945:	gap of unknown length
*	4946	9227:	contig of 4282 bp in length
*	9228	9327:	gap of unknown length
*	9328	12246:	contig of 2919 bp in length
*	12247	12346:	gap of unknown length
*	12347	15769:	contig of 3423 bp in length
*	15770	15869:	gap of unknown length

*	15870	18833:	contig of 2964 bp in length
*	18934	18933:	gap of unknown length
*	22474	22474:	contig of 3541 bp in length
*	22475	22574:	gap of unknown length
*	22575	26521:	contig of 3947 bp in length
*	26522	26621:	gap of unknown length
*	26622	29700:	contig of 3079 bp in length
*	29701	29800:	gap of unknown length
*	29801	32226:	contig of 2426 bp in length
*	32227	32326:	gap of unknown length
*	32327	34757:	contig of 2431 bp in length
*	34758	34857:	gap of unknown length
*	34858	37285:	contig of 2428 bp in length
*	37286	37385:	gap of unknown length
*	37386	40905:	contig of 3520 bp in length
*	40906	41005:	gap of unknown length
*	41006	42911:	contig of 1906 bp in length
*	42912	43011:	gap of unknown length
*	43012	45441:	contig of 2430 bp in length
*	45442	45541:	gap of unknown length
*	45542	47378:	contig of 1837 bp in length
*	47379	47478:	gap of unknown length
*	47479	50090:	contig of 2612 bp in length
*	50091	50190:	gap of unknown length
*	50191	51903:	contig of 1713 bp in length
*	51904	52003:	gap of unknown length
*	52004	55426:	contig of 3423 bp in length
*	55427	55526:	gap of unknown length
*	55527	57671:	contig of 2145 bp in length
*	57672	57771:	gap of unknown length
*	57772	59607:	contig of 1836 bp in length
*	59608	59707:	gap of unknown length
*	59708	62250:	contig of 2543 bp in length
*	62251	62350:	gap of unknown length
*	62351	64769:	contig of 2419 bp in length
*	64770	64869:	gap of unknown length
*	64870	67350:	contig of 2481 bp in length
*	67351	67450:	gap of unknown length
*	67451	69098:	contig of 1648 bp in length
*	69099	69198:	gap of unknown length
*	69199	71553:	contig of 2355 bp in length
*	71554	71653:	gap of unknown length
*	71654	72699:	contig of 1046 bp in length
*	72700	72799:	gap of unknown length
*	72800	74905:	contig of 2106 bp in length
*	74906	75005:	gap of unknown length
*	75006	76654:	contig of 1649 bp in length
*	76655	76754:	gap of unknown length
*	76755	78897:	contig of 2143 bp in length
*	78898	78997:	gap of unknown length
*	78998	80219:	contig of 1222 bp in length
*	80220	80319:	gap of unknown length
*	80320	82215:	contig of 1896 bp in length
*	82216	82315:	gap of unknown length
*	82316	83638:	contig of 1323 bp in length
*	83639	83738:	gap of unknown length
*	83739	85398:	contig of 1660 bp in length
*	85399	85498:	gap of unknown length
*	85499	86815:	contig of 1317 bp in length
*	86816	86915:	gap of unknown length
*	86916	88189:	contig of 1274 bp in length
*	88190	88289:	gap of unknown length
*	88290	89602:	contig of 1313 bp in length
*	89603	89702:	gap of unknown length
*	89703	91028:	contig of 1326 bp in length
*	91029	91128:	gap of unknown length
*	91129	93315:	contig of 2187 bp in length
*	93316	93415:	gap of unknown length
*	93416	94577:	contig of 1162 bp in length
*	94578	94677:	gap of unknown length
*	94678	96688:	contig of 2011 bp in length
*	96689	96788:	gap of unknown length
*	96789	98296:	contig of 1508 bp in length

```
   *    98297    98396: gap of unknown length
   *    99543    99542: contig of 1146 bp in length
   *    99643    99642: gap of unknown length
   *   100846   100846: contig of 1204 bp in length
   *   100947   100946: gap of unknown length
   *   102313   102312: contig of 1366 bp in length
   *   102413   102412: gap of unknown length
   *   104114   104113: contig of 1701 bp in length
   *   104214   104213: gap of unknown length
   *   105408   105407: contig of 1194 bp in length
   *   105508   105507: gap of unknown length
   *   107213   107212: contig of 1705 bp in length
   *   107313   107312: gap of unknown length
   *   109341   109340: contig of 2028 bp in length
   *   109441   109440: gap of unknown length
   *   110562   110561: contig of 1121 bp in length
   *   110662   110661: gap of unknown length
   *   111852   111851: contig of 1190 bp in length
   *   111952   111951: gap of unknown length
   *   113022   113021: contig of 1070 bp in length
   *   113122   113121: gap of unknown length
   *   114684   114683: contig of 1562 bp in length
   *   114784   114783: gap of unknown length
   *   115866   115865: contig of 1082 bp in length
   *   115966   115965: gap of unknown length
   *   117094   117093: contig of 1128 bp in length
   *   117194   117193: gap of unknown length
   *   118350   118349: contig of 1056 bp in length
   *   118250   118249: gap of unknown length
   *   119827   119826: contig of 1477 bp in length
   *   119927   119926: gap of unknown length
   *   121559   121559: contig of 1633 bp in length
```

Query Match   13.6%;   Score 68;   DB 2;   Length 143435;
Best Local Similarity 48.2%;   Pred. No. 0.052;
Matches 220; Conservative 0; Mismatches 235; Indels 1; Gaps 1;

```
Qy     18 GAGCGGCGGCAGCTGCGAGGCTCGGAGAAACAGGCGCCGCGGGCTCCGGCCCGGCCGGA  77
          ||-|||||-||-|||-|||||---|-|||-|||||||-||-||-|-|||-|-||-||||
Db  69290 GACCGGAGGTCGGCGGCCGCCTGGGCTCCGGCCCGCGGGGCGCGCGGCCCCCATGCCGCT 69349

Qy     78 CCCGGGGCCCGAGAATCATGATTGCTGCCGCGCCACCCGCCACCACGGAGCGAAGCCCAGA 137
          ||-|-|||||||-|-||||-||-||||-|-|||||||-||-|||-|-||||||||-||||
Db  69350 CACCCCGCGACGGTGAGGCGCGCCGGTGCTACGCCCCGCCACCCCCGCCCCGCGCCCCCCC 69409

Qy    138 TAGACGCGCCCGGCGGGGGCCCGGGTCCTG-GAGTCCCCGCCGCCTGCTGACCCGGCCGAGGACC 196
          ||-||||                ||||||||-||-||||||||||||||-||||||||||
Db  69410 CCGCCCCCCTCACCCGCGCCCTTGCGCCCTTGCGCCGGCCTGCTCCCGCGCTCCCGCGACGCCG 69469

Qy    197 CCACCCCGCCGCCCCGGCCCCGATGCTTGCAGTGGGGCGCCCCATGGACAGGGATTACCCGCA 256
          |||-|-||-||-||||-||||||-|||||||||-|-|||||||||||-|||-||||-|-||||
Db  69470 CGGGCCGGCGGCGCCCGCCCCCCGAGCCCGCGGCCCCACCCGGGCCCGGCCCGGCCCGCGGCC 69529

Qy    257 GCATGAACCCCCGCCCCGGCGGGGCAGCCTCCTTCTGTACAGCCCGCCCCCCTGCGAGAGCGCCCAT 316
          ||-|||-|||||-|||-||||||-|-|||-||||||-|-|||||-|||||||||||-||||-||||
Db  69530 GCCGGCCCCACCGCCCCCACGCCTCTCCCCCCCGGTCCCCGCCCCCTTCCGCGCCGCGCGGCCCCAT 69589

Qy    317 GCTGCACTGCCCCTACTGGAACACCTTCTGCGTCGCCCCATACCCTGCCATTCTCAGCGA 376
          ||||-||||-|-||||-||||-|-||-|-|||-|||-|||-|||-||-|||||||||||
Db  69590 CCCCCCCAAAACACGCCCCCANTCTCCCCCCGCCCCCCCCCCCCGCTCACAACACCCCCC 69649

Qy    377 CAGCCGCCCGGTTCATGAGCTCCGCGCCTCCTTCCTCGGGCAGCCAGCCGCCCTGCCCAGACACCAG 436
          ||-||-||-|||||-||||||||||-|-|||-|-|||-|-||||-||-||-||||-|||-||||
Db  69650 CCGCCCCCGCGCCCCATCTCTCCCCCCGCCCCCCGCGCCCCCCCCCCCGCCCCCCCAC 69709

Qy    437 CTATGCCCGTGGCCACCGCCTCCAGCTTGCCACC 472
          ||||-||||-||||||||||-|||||-|||-||-||
Db  69710 CCCCCGCCCCCGCCTCTCCCTCTCGCTCCCTCCC 69745
```

Search completed: October 10, 2002, 18:29:20
Job time : 1479.87 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                                           (without alignments)
                                           14616.225 Million cell updates/sec

Title: us-09-489-101A-14_COPY_1_500
Perfect score: 500
Sequence: 1 tcaggctgctgatctgccc.............ctcacccggcggcgcccctcc 500

Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326B293 residues

Total number of hits satisfying chosen parameters: 3595312

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :   GenEmbl:*

	Database	
1:	gb_ba:*	
2:	gb_htg:*	
3:	gb_in:*	
4:	gb_om:*	
5:	gb_ov:*	
6:	gb_pat:*	
7:	gb_ph:*	
8:	gb_pl:*	
9:	gb_pr:*	
10:	gb_ro:*	
11:	gb_sts:*	
12:	gb_sy:*	
13:	gb_un:*	
14:	gb_vi:*	
15:	em_ba:*	
16:	em_fun:*	
17:	em_hum:*	
18:	em_in:*	
19:	em_mu:*	
20:	em_om:*	
21:	em_or:*	
22:	em_ov:*	
23:	em_pat:*	
24:	em_ph:*	
25:	em_pl:*	
26:	em_ro:*	
27:	em_sts:*	
28:	em_un:*	
29:	em_vi:*	
30:	em_htg_hum:*	
31:	em_htg_inv:*	
32:	em_htg_other:*	
33:	em_htgo_inv:*	

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result		Query	%			
No.	Score	Match	Length	DB	ID	Description
---	-----	-----	------	--	--	-----------

1	500	100.0	1468	6	A81360	A81360 Sequence 5
2	500	100.0	1468	6	AX001580	AX001580 Sequence 5
3	500	100.0	1629	6	A81358	A81358 Sequence 3
4	500	100.0	1629	6	AX001578	AX001578 Sequence
5	500	100.0	1871	6	A21353	A21353 H.sapiens m
6	500	100.0	1871	6	AR014368	AR014368 Sequence
7	500	100.0	1871	6	AR083557	AR083557 Sequence
8	500	100.0	1872	9	AX201593	AX201593 Sequence
9	500	100.0	1872	9	X51985	X51985 Human LAG-3
10	500	100.0	2279	6	A81356	A81356 Sequence 1
11	500	100.0	2279	6	AX001576	AX001576 Sequence
12	291	58.2	1164	6	A21357	A21357 H.sapiens D
13	291	58.2	1164	6	AR014370	AR014370 Sequence
14	291	58.2	1164	6	AR083559	AR083559 Sequence
15	288.8	57.8	245880	2	AC079387	AC079387 Homo sapi
16	221.2	44.2	161629	2	AC087604	AC087604 Pan trogl
17	221	44.2	999	6	A21356	A21356 H.sapiens D
18	221	44.2	999	6	AR014369	AR014369 Sequence
19	221	44.2	999	6	AR083558	AR083558 Sequence
20	221	44.2	2456	9	X51984	X51984 Human LAG-3
21	221	44.2	2937	9	Y10211	Y10211 H.sapiens L
22	221	44.2	5689	6	Y13586	Y13586 Homo sapien
23	204	40.8	5689	6	AX281440	AX281440 Sequence
24	204	40.8	5689	6	AX346055	AX346055 Sequence
25	155	31.0	2003	10	MMLAG3	M.musculus
26	106.4	21.3	5689	6	HSPROLAG3	HSPROLAG3
27	106.4	21.3	5689	6	HSPTMSLAG	HSPTMSLAG
28	78	15.6	7218	6	I66494	I66494 Sequence 14
29	63	12.6	119405	2	AL669909	AL669909 Mus muscu
30	59.2	11.8	306695	2	AC079431	AC079431 Mus muscu
31	57.4	11.5	119736	2	AC092270	AC092270 Rattus no
32	57	11.4	62280	2	AC101812	AC101812 Mus muscu
33	57	11.4	64822	2	AC110270	AC110270 Mus muscu
34	56.6	11.3	258174	2	AC079429	AC079429 Mus muscu
35	55	11.0	130542	2	AC108375	AC108375 Pan trogl
36	54.8	11.0	64237	2	AC096821	AC096821 Rattus no
37	54.6	10.9	187931	2	AC073935	AC073935 Rattus no
38	54.4	10.9	164520	2	AC020738	AC020738 Homo sapi
39	54.4	10.9	183016	2	AC095502	AC095502 Rattus no
40	54.2	10.8	202173	2	AC102994	AC102994 Rattus no
41	54	10.8	65304	2	AC102172	AC102172 Mus muscu
42	54	10.8	201275	2	AL603662	AL603662 Mus muscu
43	53.8	10.8	298166	2	AC087563	AC087563 Homo sapi
44	53.8	10.8	303091	2	AC084799	AC084799 Mus muscu
45	53.6	10.7	5981	2	AC109765	AC109765 Rattus no

ALIGNMENTS

RESULT 1
A81360

LOCUS   A81360   1468 bp   DNA   linear   PAT 21-JAN-2000
DEFINITION Sequence 5 from Patent EP0900841.
ACCESSION   A81360
VERSION   A81360.1   GI:6731679
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1468)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE   LAG-3 splice variants
  JOURNAL   Patent: EP 0900841-A 5 10-MAR-1999;
            APPLIED RESEARCH SYSTEMS (AN); INST NAT SANTE RECH MED (FR)
FEATURES   Location/Qualifiers
     source   1..1468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   231 a   550 c   404 g   283 t
ORIGIN

```
Query Match      100.0%; Score 500; DB 6; Length 1468;
Best Local Similarity  100.0%; Pred. No. 2.4e-95;
Matches  500; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy    1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTCTGGATTCCGGCCTCTGGTCATC 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTCTGGATTCCGGCCTCTGGTCATC 60

Qy   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCCTCCTGGTCTCCCCTTTCTCTTAGAGAACCCCTC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCCTCCTGGTCTCCCCTTTCTCTTAGAGAACCCCTC 120

Qy  121 CACCTCCCTCTCTCTGCAGAACTTCTCCTTTTACCCCCCACCACTGCCCCCCTTTCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACCTCCCTCTCTCTGCAGAACTTCTCCTTTTACCCCCCACCACTGCCCCCCTTTCC 180

Qy  181 TTTTCTGACCTCCTTTTGGAGGGCTCAGGGGCTTCAGGCTGCCCAGCCACCCCCACCCCCGC 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TTTTCTGACCTCCTTTTGGAGGGCTCAGGGGCTTCAGGCTGCCCAGCCACCCCCACCCCCGC 240

Qy  241 CTCAGTTCCTGGGCTTGCTGTTCTTCGAGCCGCCCGCGTTCGCACCGCGTGTGGGGAGG 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CTCAGTTCCTGGGCTTGCTGTTCTTCGAGCCGCCCGCGTTCGCACCGCGTGTGGGGAGG 300

Qy  301 AGCCAGGGGGCTGAGGTCCCGGGTCCCGGGCCCTTTGGGTGGTGGGCTCACCGGTCCCT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AGCCAGGGGGCTGAGGTCCCGGGTCCCGGGCCCTTTGGGTGGTGGGCTCACCGGTCCCT 360

Qy  361 GCAGCCCCACAAATCCCCTCCAGGATCTCAGCCTTCTCGGAAGACAGGGGTCACTTGGC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCAGCCCCACAAATCCCCTCCAGGATCTCAGCCTTCTCGGAAGACAGGGGTCACTTGGC 420

Qy  421 AGCATCAGCCAGACAGTGGCCCCGCCGCCGCCGCCCGGGCGCCCATCCCCTGGTCACTTGGC 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGCATCAGCCAGACAGTGGCCCCGCCGCCGCCGCCCGGGCGCCCATCCCCTGGTCACTTGGC 480

Qy  481 CTCACCCGGCGGGGGCGCGCCCTCC 500
        |||||||||||||||||||||||||
Db  481 CTCACCCGGCGGGGGCGCGCCCTCC 500
```

RESULT 3

```
LOCUS       A81358          1629 bp    DNA     linear   PAT 21-JAN-2000
DEFINITION  Sequence 3 from Patent EP0900841.
ACCESSION   A81358
VERSION     A81358.1  GI:6731678
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE     LAG-3 splice variants
  JOURNAL   Patent: EP 0900841-A 3 10-MAR-1999;
            APPLIED RESEARCH SYSTEMS (AN); INST NAT SANTE RECH MED (FR)
FEATURES             Location/Qualifiers
     source          1..1629
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      252 a    605 c    452 g    320 t
ORIGIN
```

```
Query Match      100.0%; Score 500; DB 6; Length 1629;
Best Local Similarity  100.0%; Pred. No. 2.3e-95;
Matches  500; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy    1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTCTGGATTCCGGCCTCTGGTCATC 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTCTGGATTCCGGCCTCTGGTCATC 60

Qy   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCCTCCTGGTCTCCCCTTTCTCTTAGAGAACCCCTC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCTCCCCACCCTCTCTCCAAGGCCCCTCCTCCTGGTCTCCCCTTTCTCTTAGAGAACCCCTC 120

Qy  121 CACCTCCCTCTCTCTGCAGAACTTCTCCTTTTACCCCCCACCACTGCCCCCCTTTCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACCTCCCTCTCTCTGCAGAACTTCTCCTTTTACCCCCCACCACTGCCCCCCTTTCC 180

Qy  181 TTTTCTGACCTCCTTTTGGAGGGCTCAGGGGCTTCAGGCTGCCCAGCCACCCCCACCCCCGC 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 4
LOCUS       AX001578              1629 bp    DNA        linear    PAT 10-MAR-2000
DEFINITION  Sequence 3 from Patent WO9858059.
ACCESSION   AX001578
VERSION     AX001578.1  GI:7241709
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE     LAG-3 SPLICE VARIANTS
  JOURNAL   Patent: WO 9858059-A 3 23-DEC-1998;
            INST NAT SANTE RECH MED (FR); ROUSSY INST GUSTAVE (FR)
FEATURES             Location/Qualifiers
     source          1..1629
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      252 a    605 c    452 g    320 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 1629;
Best Local Similarity   100.0%;
Matches  500;           Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGTCATC                                        60
          ||||||||||||||||||||||||||||||||||||||
Db      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGTCATC                                        60

Qy     61 CCTCCCCACCCTCTCCTGGTCTCCCTCTTCTCTCTGAATTCCGGCCTCTGGTCATC                       120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTCCCCACCCTCTCCTGGTCTCCCTCTTCTCTCTGAATTCCGGCCTCTGGTCATC                       120

Qy    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC                           180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC                           180

Qy    181 TTTTCTGACCTCCTTTTGGAGGGCTCAGCGCTCCAGTGAAGCCTCACG                               240
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTTCTGACCTCCTTTTGGAGGGCTCAGCGCTCCAGTGAAGCCTCACG                               240

Qy    241 CTCAGTTCCTGCTGTCTGTTTCTGCAGCCGCGCTGCCCAGGACCATAGGAGAGATGTGGGAG                  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCAGTTCCTGCTGTCTGTTTCTGCAGCCGCGCTGCCCAGGACCATAGGAGAGATGTGGGAG                  300

Qy    301 AGCCAGGGGGCTGAGGTCCCGGGTGGGTGTGGGGCCCAGGAGAGGGGGCTCCTGCCCAGCTCCCCT             360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCAGGGGGCTGAGGTCCCGGGTGGGTGTGGGGCCCAGGAGAGGGGGCTCCTGCCCAGCTCCCCT             360

Qy    361 GCAGCCCCCACCAAATCCCCTCCAGGAATCTCAGCCTTTCTGCCGAAGAGAGCAGGGGTCACTTGGC            420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCCCCCACCAAATCCCCTCCAGGAATCTCAGCCTTTCTGCCGAAGAGAGCAGGGGTCACTTGGC            420

Qy    421 AGCATCAGCCAGACAGAGTGGCCGCGCTGCCGCGCCATCCCCTGCCCCGGGCC             480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCATCAGCCAGACAGAGTGGCCGCGCTGCCGCGCCATCCCCTGCCCCGGGCC             480

Qy    481 CTCACCCCGGCGGCGGCCCTCC             500
          ||||||||||||||||||||||
Db    481 CTCACCCCGGCGGCGGCCCTCC             500
```

```
RESULT 5
LOCUS       A21353                1871 bp    mRNA       linear    PAT 05-JUL-1994
DEFINITION  H.sapiens mRNA for protein from lymphocytes.
ACCESSION   A21353
VERSION     A21353.1  GI:579595
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1871)
  AUTHORS   .
  TITLE     PROTEINS PRODUCED BY HUMAN LYMPHOCYTES, DNA SEQUENCE CODING THESE
            PROTEINS, AND PHARMACEUTICAL AND BIOLOGICAL USES THEREOF
  JOURNAL   Patent: WO 9110682-A 1 25-JUL-1991;
FEATURES             Location/Qualifiers
     source          1..1871
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             231..1727
                     /codon_start=1
                     /product="lymphocyte protein"
                     /protein_id="CAA01547.1"
                     /db_xref="GI:579596"
                     /db_xref="SWISS-PROT:P18627"
                     /translation="MWEAQFLGLLFLQPLWVAPVKPLQPGAEVPVVWNAQEGAPAQLPC
                     SPTIPLQDLSLLRRAGVTWQHQPDSGPPAAAPGHPLAPGPHPAAPASSWGPRPRRYTVL
                     SVGPGGLRSGRLPLQPRVQLDERGRQRGDFSLWLRPARRADAGEYRAAVHLRDRALSC
                     RLRLRLGQASMTASPPGSLRASDWVILNCSFSRPDRPASVHWFRNRGGRVPVRESPH
                     HHLAESFLFLPQVSPMDSGPWGCILTYRDGFNVSIMYNLTVLGLEPPTPLTVYAGAGS
                     RVGLPCRLPAGVGTRSFLTAKWTPPGGGVDLLVTGDNGDFTLRLEDVSQAQAGTYTCH
                     IHLQEQLNATVTLAIITVTPKSFGSPGSLGKLLCEVTPVSGQERFVWNSSLDTPSQRS
                     FSGPWLEAQEAQLLSQPWQCQLYQGERLLGAAVYFTELSSPGAQRSRAPGALPAGHL
                     LLFLLGVLSLLLLVTGAFGFHLWRRQWRPRRFSALEQGIHPRRLRAR"
BASE COUNT      300 a    672 c    523 g    376 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 1871;
Best Local Similarity   100.0%;
Matches  500;           Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGTCATC                                        60
          ||||||||||||||||||||||||||||||||||||||
Db      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCTCTGGTCATC                                        60

Qy     61 CCTCCCCACCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC                         120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTCCCCACCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC                         120

Qy    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC                            180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC                            180

Qy    181 TTTTCTGACCTCCTTTTGGAGGGCTCAGCGCTCCAGTGAAGCCTCACG                                240
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTTCTGACCTCCTTTTGGAGGGCTCAGCGCTCCAGTGAAGCCTCACG                                240

Qy    241 CTCAGTTCCTGGCTTGCTGTTTCTGCACCGCGCTTGGGTGGGTGGGGCTCCCAGTGAAGCCTCTCC             300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCAGTTCCTGGCTTGCTGTTTCTGCACCGCGCTTGGGTGGGTGGGGCTCCCAGTGAAGCCTCTCC             300
```

```
Qy    361 GCAGCCCCCACCAAATCCCCTCCAGGAATCTCAGCCTTTCTGCCGAAGAGAGCAGGGGTCACTTGGC            420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCCCCCACCAAATCCCCTCCAGGAATCTCAGCCTTTCTGCCGAAGAGAGCAGGGGTCACTTGGC            420

Qy    421 AGCATCAGCCAGACAGAGTGGCCGCGCTGCCGCGCCATCCCCTGCCCCGGGCC                          480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCATCAGCCAGACAGAGTGGCCGCGCTGCCGCGCCATCCCCTGCCCCGGGCC                          480

Qy    481 CTCACCCCGGCGGCGGCCCTCC             500
          ||||||||||||||||||||||
Db    481 CTCACCCCGGCGGCGGCCCTCC             500
```

RESULT 6

```
LOCUS       AR014368                1871 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 1 from patent US 5773578.
ACCESSION   AR014368
VERSION     AR014368.1  GI:3971822
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1871)
  AUTHORS   Hercend,T. and Triebel,F.
  TITLE     Proteins produced by human lymphocytes, DNA sequence encoding these
            proteins and their pharmaceutical and biological use
  JOURNAL   Patent: US 5773578-A 1 30-JUN-1998;
FEATURES             Location/Qualifiers
     source          1..1871
                     /organism="unknown"
BASE COUNT      300 a    672 c    523 g    376 t
ORIGIN
```

Query Match   100.0%; Score 500; DB 6; Length 1871;
Best Local Similarity   100.0%;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCGCCCTGGTCATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCGCCCTGGTCATC 60

Qy     61 CCTCCCCACCCCTCTCTCCAAAGGCCCTCTCCTTGGTGTCCCTTCTTCTTCTAGAACCCTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTCCCCACCCCTCTCTCCAAAGGCCCTCTCCTTGGTGTCCCTTCTTCTTCTAGAACCCTC 120

Qy    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCCACCCCCCACCACTGCCCCCCTTTCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCCACCCCCCACCACTGCCCCCCTTTCC 180

Qy    181 TTTTTCTGACCTCCTTTTTGGAGGGCTCAGCGCCTGCCCAGACCATAGGAGAGAGTGGGAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTTTCTGACCTCCTTTTTGGAGGGCTCAGCGCCTGCCCAGACCATAGGAGAGAGTGGGAGG 240

Qy    241 CTCAGTTCCTGGGCTTGGCTTCTGCACCGCGTTGTGGGCTCCAGCGAGCCTCTC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCAGTTCCTGGGCTTGGCTTCTGCACCGCGTTGTGGGCTCCAGCGAGCCTCTC 300

Qy    301 AGCCAGGGGGCTGAGGTGGTCCCGGGTGCGTGGGGCCCCAGGGGGCTCCTCCTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCAGGGGGCTGAGGTGGTCCCGGGTGCGTGGGGCCCCAGGGGGCTCCTCCTG 360

Qy    361 GCAGCCCCACAATCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGAGCAGGGGGTCACTTGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCCCCACAATCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGAGCAGGGGGTCACTTGGC 420

Qy    421 AGCATCAGCCAGACAGTGGCCCGCGCCGCCGCCGCCGCTGGCCGCCATCCCCTGGCCCCGGCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCATCAGCCAGACAGTGGCCCGCGCCGCCGCCGCCGCTGGCCGCCATCCCCTGGCCCCGGCC 480

Qy    481 CTCACCCGGCGGCGGCGCCCTCC 500
          ||||||||||||||||||||||
Db    481 CTCACCCGGCGGCGGCGCCCTCC 500
```

RESULT 7

```
LOCUS       AR083557                1871 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 1 from patent US 5976877.
ACCESSION   AR083557
VERSION     AR083557.1  GI:10010330
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1871)
  AUTHORS   Hercend,T. and Triebel,F.
  TITLE     Proteins produced by human lymphocytes DNA sequence encoding these
            proteins and their pharmaceutical and biological uses
  JOURNAL   Patent: US 5976877-A 1 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1..1871
                     /organism="unknown"
BASE COUNT      300 a    672 c    523 g    376 t
ORIGIN
```

Query Match   100.0%; Score 500; DB 6; Length 1871;
Best Local Similarity   100.0%;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCGCCCTGGTCATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCGCCCTGGTCATC 60

Qy     61 CCTCCCCACCCCTCTCTCCAAAGGCCCTCTCCTTGGTGTCCCTTCTTCTTCTAGAACCCTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTCCCCACCCCTCTCTCCAAAGGCCCTCTCCTTGGTGTCCCTTCTTCTTCTAGAACCCTC 120

Qy    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCCACCCCCCACCACTGCCCCCCTTTCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACCTCCCTCTCTGCAGAACTTCTCCTTTACCCCCCACCCCCCACCACTGCCCCCCTTTCC 180

Qy    181 TTTTTCTGACCTCCTTTTTGGAGGGCTCAGCGCCTGCCCAGACCATAGGAGAGAGTGGGAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTTTCTGACCTCCTTTTTGGAGGGCTCAGCGCCTGCCCAGACCATAGGAGAGAGTGGGAGG 240

Qy    241 CTCAGTTCCTGGGCTTGGCTTCTGCACCGCGTTGTGGGCTCCAGCGAGCCTCTC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTCAGTTCCTGGGCTTGGCTTCTGCACCGCGTTGTGGGCTCCAGCGAGCCTCTC 300

Qy    301 AGCCAGGGGGCTGAGGTGGTCCCGGGTGCGTGGGGCCCCAGGGGGCTCCTCCTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCAGGGGGCTGAGGTGGTCCCGGGTGCGTGGGGCCCCAGGGGGCTCCTCCTG 360

Qy    361 GCAGCCCCACAATCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGAGCAGGGGGTCACTTGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGCCCCACAATCCCCTCCAGGATCTCAGCCTTCTGCGAAGAGAGCAGGGGGTCACTTGGC 420

Qy    421 AGCATCAGCCAGACAGTGGCCCGCGCCGCCGCCGCCGCTGGCCGCCATCCCCTGGCCCCGGCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCATCAGCCAGACAGTGGCCCGCGCCGCCGCCGCCGCTGGCCGCCATCCCCTGGCCCCGGCC 480

Qy    481 CTCACCCGGCGGCGGCGCCCTCC 500
          ||||||||||||||||||||||
Db    481 CTCACCCGGCGGCGGCGCCCTCC 500
```

RESULT 8

AX201593

```
LOCUS           AX201593      1872 bp    DNA      linear    PAT 30-AUG-2001
DEFINITION      Sequence 14 from Patent WO0153349.
ACCESSION       AX201593
VERSION         AX201593.1  GI:15391442
KEYWORDS        .
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 1872)
  AUTHORS       Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
                Chen,Y.T.
  TITLE         Small cell lung cancer associated antigens and uses therefor
  JOURNAL       Patent: WO 0153349-A1 4 26-JUL-2001;
                LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
                SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
                INC. (US)
FEATURES             Location/Qualifiers
     source          1..1872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      300 a    673 c    523 g    376 t
ORIGIN
Query Match      100.0%;  Score 500;  DB 6;  Length 1872;
Best Local Similarity  100.0%;  Pred. No. 2.3e-95;
Matches 500;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60

Qy       61 CCTCCCCACCCTCTCTCCAAGGCCCTCTCCTGGTCTCCCCTTCTTCTTCTAGAACCCCTC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CCTCCCCACCCTCTCTCCAAGGCCCTCTCCTGGTCTCCCCTTCTTCTTCTAGAACCCCTC 120

Qy      121 CACCTTCCCTCTGCAGAACTTCTCCTTTACCCCCCACCCACTCACTGCCCCCTTTCC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CACCTTCCCTCTGCAGAACTTCTCCTTTACCCCCCACCCACTCACTGCCCCCTTTCC 180

Qy      181 TTTTCTGACCTCCTTTTGGAGGGGCTCAGGGGTGCCCAGACCATAGGAGAGAATGTGGGAGG 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TTTTCTGACCTCCTTTTGGAGGGGCTCAGGCGTGCCCAGACCATAGGAGAGAATGTGGGAGG 240

Qy      241 CTCAGTTCCTGGGCTTGCTGTTTCTGCGAGCCCCGCCTCCAGTGAAGCCTCTCC 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CTCAGTTCCTGGGCTTGCTGTTTCTGCGAGCCCCGCCTCCAGTGAAGCCTCTCC 300

Qy      301 AGCCAGGGGCTGAGTCCCGGTGGTGGTCCCCAGAGGGGGCTCGCCCAGCCTCCCCT 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AGCCAGGGGCTGAGTCCCGGTGGTGGTCCCCAGAGGGGGCTCGCCCAGCCTCCCCT 360

Qy      361 GCAGCCCCACAATCCCCTCCCTTCCAGGATCTCAGCCTTCTGCGAAGCAGCAGGGGTCACTTGGC 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GCAGCCCCACAATCCCCTCCCTTCCAGGATCTCAGCCTTCTGCGAAGCAGCAGGGGTCACTTGGC 420

Qy      421 AGCATCAGCCAGACAGTGGCCCGCGCCGCCCGCCCCGGGC 480
            ||||||||||||||||||||||||||||||||||||||||
Db      421 AGCATCAGCCAGACAGTGGCCCGCGCCGCCCGCCCCGGGC 480

Qy      481 CTCACCCGGCGCGGCGCCCTCC 500
            ||||||||||||||||||||||
Db      481 CTCACCCGGCGCGGCGCCCTCC 500
```

```
RESULT 9
HSLAG3
LOCUS         HSLAG3         1872 bp    mRNA      linear    PRI 12-SEP-2001
DEFINITION    Human LAG-3 mRNA for CD4-related protein involved in lymphocyte
              activation.
ACCESSION     X51985
VERSION       X51985.3  GI:15617340
KEYWORDS      cell surface glycoprotein; immune response; immunoglobulin
```

```
                superfamily; transmembrane protein.
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 1872)
  AUTHORS       Triebel,F., Jitsukawa,S., Baixeras,E., Roman-Roman,S., Genevee,C.,
                Viegas-Pequignot,E. and Hercend,T.
  TITLE         LAG-3, a novel lymphocyte activation gene closely related to CD4
  JOURNAL       The Journal of experimental medicine. 171 (5), 1393-1405 (1990)
  MEDLINE       90237736
REFERENCE       2  (bases 1 to 1872)
  AUTHORS       Triebel,F.
  TITLE         Direct Submission
  JOURNAL       Submitted (26-FEB-1990) Triebel F., Laboratoire d'Immunologie
                Cellulaire U333, Institut Gustave Roussy rue Camille Desmoulins,
                94805 Villejuif, France
REMARK          Revised by [3]
REFERENCE       3  (bases 1 to 1872)
  AUTHORS       Triebel,F.
  TITLE         Direct Submission
  JOURNAL       Submitted (12-AUG-1996) Triebel F., Laboratoire d'Immunologie
                Cellulaire U333, Institut Gustave Roussy rue Camille Desmoulins,
                94805 Villejuif, France
REMARK          Revised by [4]
REFERENCE       4  (bases 1 to 1872)
  AUTHORS       Triebel,F.
  TITLE         Direct Submission
  JOURNAL       Submitted (21-NOV-2000) Triebel F., Laboratoire d'Immunologie
                Cellulaire U333, Institut Gustave Roussy rue Camille Desmoulins,
                94805 Villejuif, France
COMMENT         On Sep 13, 2001 this sequence version replaced gi:11558021.
                Data kindly reviewed (08-OCT-1990) by Triebel F.
FEATURES             Location/Qualifiers
     source          1..1872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MB-F(5)"
     sig_peptide     231..296
                     /note="signal peptide"
     CDS             231..1808
                     /codon_start=1
                     /product="LAG-3 protein precursor"
                     /db_xref="GI:15617341"
                     /db_xref="SWISS-PROT:P18627"
                     /translation="MWEAQFLGLLFLQPLWVAPVKPLQPGAEVPVWAQEGAPAQLPC
                     SPTIPLQDISLLRRAGVTWQHQPDSGPPAAAPGHPLAPGPHPAAPSSWGPPPRPRYTVL
                     SVGPGGLRSRGRLPLQPRVQLDRGRQRDFSLWLRPARRADAGEYRAAVHLRDALSC
                     RLRLRLQQASMPASPPGSLRASDWVILNCSFSRPDRPASVHWFRNRGQGRVPVRESPH
                     HHLAESFLPLPQVSPMDSGPWGCILTYRDGFNVSIMYNLTVLLGLEPPTPLTVVAGAGS
                     RVGLPCRLPAGVGTRSFLTAKWTPPGGGPDLLVTGDNGDFTLRLEDVSQAAQGTYTCH
                     IHLQEQQLNATVTLAIITVTPKSFGSPGSLGKLLCEVTPVSGQERFVWSSLDTPSQRS
                     FSGPWLEAQEAQLLSQPWQCQLYQGERLLGAAVYFTELSSPGAQRSRAPGALPAGHL
                     LFLTLGVILSLLLLVTGAFGFHLWRROWPRRSALEQGIHPPQAQSKIEELEQPEP
                     EPEPEPEPEPEPEQL"
     mat_peptide     297..1805
                     /product="LAG-3 protein"
BASE COUNT      300 a    673 c    523 g    376 t
ORIGIN
Query Match      100.0%;  Score 500;  DB 9;  Length 1872;
Best Local Similarity  100.0%;  Pred. No. 2.3e-95;
Matches 500;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTTCCTCTGGATTCCGGCCTCTGGTCATC 60

Qy       61 CCTCCCCACCCTCTCTCCAAGGCCCTCTCCTGGTCTCCCCTTCTTCTTCTAGAACCCCTC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CCTCCCCACCCTCTCTCCAAGGCCCTCTCCTGGTCTCCCCTTCTTCTTCTAGAACCCCTC 120
```

```
RESULT 10
LOCUS       A81356          2279 bp    DNA        linear   PAT 21-JAN-2000
DEFINITION  A81356 Sequence 1 from Patent EP0900841.
ACCESSION   A81356
VERSION     A81356.1  GI:6731677
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2279)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE     LAG-3 Splice variants
  JOURNAL   Patent: EP 090084l-A1 10-MAR-1999;
            APPLIED RESEARCH SYSTEMS (AN); INST NAT SANTE RECH MED (FR)
FEATURES             Location/Qualifiers
     source          1..2279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      360 a    832 c    589 g    498 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 2279;
Best Local Similarity   100.0%;  Pred. No. 2.2e-95;
Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTCTCCCTTCCTTGGTCATC    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTCTCCCTTCCTTGGTCATC    60

Qy    61  CCTCCCACCCTCTCTCCAAGGCCCCTCCTCCTTGGTCTCCTTCTTCTAGAAC   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CCTCCCACCCTCTCTCCAAGGCCCCTCCTCCTTGGTCTCCTTCTTCTAGAAC   120

Qy   121  CACCTCCCTCTCTGCAGAAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CACCTCCCTCTCTGCAGAAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC  180

Qy   181  TTTTTCTGACCTCCTTTTTGGAGGGGCTCAGCGCGCTCCCTTACCCTCGTCATC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  TTTTTCTGACCTCCTTTTTGGAGGGGCTCAGCGCGCTCCCTTACCCTCGTCATC 240

Qy   241  CTCAGTTCCTGGGCTTGTGTTTCTGCAGTGAAGCCTCTCC              300
          |||||||||||||||||||||||||||||||||||||||
Db   241  CTCAGTTCCTGGGCTTGTGTTTCTGCAGTGAAGCCTCTCC              300

Qy   301  AGCCAGGGGGCTGAGGTCCCGGTTGGTGGGGCTTCCAGCGCCTCC          360
          |||||||||||||||||||||||||||||||||||||||||||||
Db   301  AGCCAGGGGGCTGAGGTCCCGGTTGGTGGGGCTTCCAGCGCCTCC          360

Qy   361  GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCGCCTTCTCGGAAGAGCAGGG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCGCCTTCTCGGAAGAGCAGGG 420

Qy   421  AGCATCAGCCAGACAGTGGCCCCGCCCGCGGCCATCCCTGGCCCCCGGCC     480
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  AGCATCAGCCAGACAGTGGCCCCGCCCGCGGCCATCCCTGGCCCCCGGCC     480

Qy   481  CTCACCCGGCGGCGGCCCTCC                                 500
          |||||||||||||||||||||
Db   481  CTCACCCGGCGGCGGCCCTCC                                 500
```

```
RESULT 11
LOCUS       AX001576        2279 bp    DNA        linear   PAT 10-MAR-2000
DEFINITION  AX001576 Sequence 1 from Patent WO9858059.
ACCESSION   AX001576
VERSION     AX001576.1  GI:7241708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2279)
  AUTHORS   Triebel,F. and Mastrangeli,R.
  TITLE     LAG-3 SPLICE VARIANTS
  JOURNAL   Patent: WO 9858059-A1 23-DEC-1998;
            INST NAT SANTE RECH MED (FR); ROUSSY INST GUSTAVE (FR)
FEATURES             Location/Qualifiers
     source          1..2279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      360 a    832 c    589 g    498 t
ORIGIN

Query Match             100.0%;  Score 500;  DB 6;  Length 2279;
Best Local Similarity   100.0%;  Pred. No. 2.2e-95;
Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTCTCCCTTCCTTGGTCATC    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  TCAGGCTGCCTGATCTGCCCAGCTTTCCAGCTTCTCCCTTCCTTGGTCATC    60

Qy    61  CCTCCCACCCTCTCTCCAAGGCCCCTCCTCCTTGGTCTCCTTCTTCTAGAAC   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CCTCCCACCCTCTCTCCAAGGCCCCTCCTCCTTGGTCTCCTTCTTCTAGAAC   120

Qy   121  CACCTCCCTCTCTGCAGAAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CACCTCCCTCTCTGCAGAAACTTCTCCTTTACCCCCACCACTGCCCCCTTTCC  180

Qy   181  TTTTTCTGACCTCCTTTTTGGAGGGGCTCAGCGCGCTCCCTTACCCTCGTCATC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  TTTTTCTGACCTCCTTTTTGGAGGGGCTCAGCGCGCTCCCTTACCCTCGTCATC 240

Qy   241  CTCAGTTCCTGGGCTTGTGTTTCTGCAGTGAAGCCTCTCC              300
          |||||||||||||||||||||||||||||||||||||||
Db   241  CTCAGTTCCTGGGCTTGTGTTTCTGCAGTGAAGCCTCTCC              300

Qy   301  AGCCAGGGGGCTGAGGTCCCGGTTGGTGGGGCTTCCAGCGCCTCC          360
          |||||||||||||||||||||||||||||||||||||||||||||
Db   301  AGCCAGGGGGCTGAGGTCCCGGTTGGTGGGGCTTCCAGCGCCTCC          360

Qy   361  GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCGCCTTCTCGGAAGAGCAGGG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GCAGCCCCACAAATCCCCCTCCAGGATCTCAGCGCCTTCTCGGAAGAGCAGGG 420

Qy   421  AGCATCAGCCAGACAGTGGCCCCGCCCGCGGCCATCCCTGGCCCCCGGCC     480
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  AGCATCAGCCAGACAGTGGCCCCGCCCGCGGCCATCCCTGGCCCCCGGCC     480

Qy   481  CTCACCCGGCGGCGGCCCTCC                                 500
          |||||||||||||||||||||
Db   481  CTCACCCGGCGGCGGCCCTCC                                 500
```

RESULT 12

```
LOCUS       A21357                   1164 bp   DNA      linear    PAT 05-JUL-1994
DEFINITION  H.sapiens DNA sequence.
ACCESSION   A21357
VERSION     A21357.1  GI:579598
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1164)
  AUTHORS
  TITLE     PROTEINS PRODUCED BY HUMAN LYMPHOCYTES, DNA SEQUENCE CODING THESE
            PROTEINS, AND PHARMACEUTICAL AND BIOLOGICAL USES THEREOF
  JOURNAL   Patent: WO 9110682-A 5 25-JUL-1991;
FEATURES             Location/Qualifiers
  source    1..1164
            /db_xref="taxon:9606"
            /organism="Homo sapiens"
BASE COUNT    182 a    426 c    340 g    216 t
ORIGIN
```

Query Match 58.2%; Score 291; DB 6; Length 1164;
Best Local Similarity 100.0%; Pred. No. 2e-51;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  210 GCTGCCCAGACCAATAGGAGAGAAGATGTGGGAGGCTCAGTTCCTGGGGCTTGCTGTTTCTGCAG 269
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCTGCCCAGACCAATAGGAGAGAAGATGTGGGAGGCTCAGTTCCTGGGGCTTGCTGTTTCTGCAG  60

Qy  270 CCGCTTTGGGTGGTCCTCCAGTGAAGCCTCTCCAGCCACGCAGGGGTCCGGTGGTGGTGTGG 329
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTTTGGGTGGTCCTCCAGTGAAGCCTCTCCAGCCACGCAGGGGTCCGGTGGTGGTGTGG 120

Qy  330 GCCCAGGAGGGGGGGGCTCTGCCCCAGTCCCTGCCCCACAAATCCCCTCCAGGAGATCTC 389
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GCCCAGGAGGGGGGGGCTCTGCCCCAGTCCCTGCCCCACAAATCCCCTCCAGGAGATCTC 180

Qy  390 AGCCTTCTCGAAGACAGCAGGGGTCACTTGGCCAGCATCAGCCAGACAGTGGCCCGCCCGCT 449
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGCCTTCTCGAAGACAGCAGGGGTCACTTGGCCAGCATCAGCCAGACAGTGGCCCGCCCGCT 240

Qy  450 GCCGCCCCCGGCCATCCCCTGGCCCCCGGCCCCGCCGGCGGCGGCGCCCCCGGCGGCGGCCCCTCC 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GCCGCCCCCGGCCATCCCCTGGCCCCCGGCCCCGCCGGCGGCGGCGCCCCCGGCGGCGGCCCCTCC 291
```

RESULT 13

```
LOCUS       AR014370                 1164 bp   DNA      linear    PAT 05-DEC-1998
DEFINITION  Sequence 5 from patent US 5773578.
ACCESSION   AR014370
VERSION     AR014370.1  GI:3971824
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1164)
  AUTHORS   Hercend,T. and Triebel,F.
  TITLE     Proteins produced by human lymphocytes, DNA sequence encoding these
            proteins and their pharmaceutical and biological use
  JOURNAL   Patent: US 5773578-A 5 30-JUN-1998;
FEATURES             Location/Qualifiers
  source    1..1164
            /organism="unknown"
BASE COUNT    182 a    426 c    340 g    216 t
ORIGIN
```

Query Match 58.2%; Score 291; DB 6; Length 1164;
Best Local Similarity 100.0%; Pred. No. 2e-51;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  210 GCTGCCCAGACCAATAGGAGAGAAGATGTGGGAGGCTCAGTTCCTGGGGCTTGCTGTTTCTGCAG 269
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCTGCCCAGACCAATAGGAGAGAAGATGTGGGAGGCTCAGTTCCTGGGGCTTGCTGTTTCTGCAG  60

Qy  270 CCGCTTTGGGTGGTCCTCCAGTGAAGCCTCTCCAGCCACGCAGGGGTCCGGTGGTGGTGTGG 329
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTTTGGGTGGTCCTCCAGTGAAGCCTCTCCAGCCACGCAGGGGTCCGGTGGTGGTGTGG 120

Qy  330 GCCCAGGAGGGGGGGGCTCTGCCCCAGTCCCTGCCCCACAAATCCCCTCCAGGAGATCTC 389
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GCCCAGGAGGGGGGGGCTCTGCCCCAGTCCCTGCCCCACAAATCCCCTCCAGGAGATCTC 180

Qy  390 AGCCTTCTCGAAGACAGCAGGGGTCACTTGGCCAGCATCAGCCAGACAGTGGCCCGCCCGCT 449
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGCCTTCTCGAAGACAGCAGGGGTCACTTGGCCAGCATCAGCCAGACAGTGGCCCGCCCGCT 240
```

RESULT 14

```
LOCUS       AR083559                 1164 bp   DNA      linear    PAT 01-SEP-2000
DEFINITION  Sequence 5 from patent US 5976877.
ACCESSION   AR083559
VERSION     AR083559.1  GI:10010332
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1164)
  AUTHORS   Hercend,T. and Triebel,F.
  TITLE     Proteins produced by human lymphocytes DNA sequence encoding these
            proteins and their pharmaceutical and biological uses
  JOURNAL   Patent: US 5976877-A 5 02-NOV-1999;
FEATURES             Location/Qualifiers
  source    1..1164
            /organism="unknown"
BASE COUNT    182 a    426 c    340 g    216 t
ORIGIN
```

Query Match 58.2%; Score 291; DB 6; Length 1164;
Best Local Similarity 100.0%; Pred. No. 2e-51;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  210 GCTGCCCAGACCAATAGGAGAGAAGATGTGGGAGGCTCAGTTCCTGGGGCTTGCTGTTTCTGCAG 269
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCTGCCCAGACCAATAGGAGAGAAGATGTGGGAGGCTCAGTTCCTGGGGCTTGCTGTTTCTGCAG  60

Qy  270 CCGCTTTGGGTGGTCCTCCAGTGAAGCCTCTCCAGCCACGCAGGGGTCCGGTGGTGGTGTGG 329
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGCTTTGGGTGGTCCTCCAGTGAAGCCTCTCCAGCCACGCAGGGGTCCGGTGGTGGTGTGG 120

Qy  330 GCCCAGGAGGGGGGGGCTCTGCCCCAGTCCCTGCCCCACAAATCCCCTCCAGGAGATCTC 389
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GCCCAGGAGGGGGGGGCTCTGCCCCAGTCCCTGCCCCACAAATCCCCTCCAGGAGATCTC 180

Qy  390 AGCCTTCTCGAAGACAGCAGGGGTCACTTGGCCAGCATCAGCCAGACAGTGGCCCGCCCGCT 449
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGCCTTCTCGAAGACAGCAGGGGTCACTTGGCCAGCATCAGCCAGACAGTGGCCCGCCCGCT 240

Qy  450 GCCGCCCCCGGCCATCCCCTGGCCCCCGGCCCCGCCGGCGGCGGCGCCCCCGGCGGCGGCCCCTCC 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GCCGCCCCCGGCCATCCCCTGGCCCCCGGCCCCGCCGGCGGCGGCGCCCCCGGCGGCGGCCCCTCC 291

Qy  361 GCAGCCCCACCAAATCCCCTCCAGGATCTCAGCCTTCTCGGAAGAGACAGCAGGGGTCACTTGGC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCAGCCCCACCAAATCCCCTCCAGGATCTCAGCCTTCTCGGAAGAGACAGCAGGGGTCACTTGGC 420

Qy  421 AGCATCAGCCAGACAGTGGCGCCGCCCGCCGCCGCCCCCGGCCATCCCCTGGCCCCCGGCC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGCATCAGCCAGACAGTGGCGCCGCCCGCCGCCGCCCCCGGCCATCCCCTGGCCCCCGGCC 480

Qy  481 CTCACCCGGCGGCGGCCCCTCC 500
        ||||||||||||||||||||||
Db  481 CTCACCCGGCGGCGGCCCCTCC 500
```

```
QY     450  GCCCGCCCCGGCCCATCCCTGGCCCCGCGGCGGCGGCGCCCTCC  500
            ||||||||||||||||||||||||||||||||||||||||||||
Db     241  GCCGGCCCCGGCCCATCCCTGGCCCCGCGGCGGCGGCGCCCTCC  291
```

```
RESULT 15
LOCUS       AC079387/c           245880 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 12 clone RP4-761J14, RP11-433J6, ***
            SEQUENCING IN PROGRESS ***, 42 unordered pieces.
ACCESSION   AC079387 AC006087 AC007438 AC006086
VERSION     AC079387.1  GI:9961230
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245880)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245880)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:4589937 gi:4926836
            gi:4589938.
            --------------    Genome Center
                    Center: Baylor College of Medicine
                    Center code: BCM
                    Web site: http://www.hgsc.bcm.tmc.edu/
                    Contact: hgsc-help@bcm.tmc.edu
            --------------    Project Information
                    Center project name: J-31
                    Center clone name: RP4-761J14, RP11-433J6
            --------------    Summary Statistics
                    Sequencing vector: M13; L08821
                    Chemistry: Dye-primer BodyJ: 92% of reads
                    Chemistry: Dye-terminator Big Dye: 7% of reads
                    Assembly program: Phrap; version 0.990329
                    Consensus quality: 214565 bases at least Q40
                    Consensus quality: 230283 bases at least Q30
                    Consensus quality: 240000 bases at least Q20
                    Estimated insert size: 29807; agarose-fp estimation
                    Quality coverage: 0x in Q20 bases; agarose-fp estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *    (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *    consists of 42 contigs. The true order of the pieces
            *    is not known and their order in this sequence record is
            *    arbitrary. Gaps between the contigs are represented as
            *    runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            *    as soon as it is available and the accession number will
            *    be preserved.
            *        1    27068: contig of 27068 bp in length
            *    27069    27168: gap of unknown length
            *    27169    51197: contig of 24029 bp in length
            *    51198    51297: gap of unknown length
            *    51298    89527: contig of 38230 bp in length
            *    89528    89627: gap of unknown length
            *    89628   108061: contig of 18434 bp in length
            *   108062   108161: gap of unknown length
            *   108162   129355: contig of 21194 bp in length
            *   129356   129455: gap of unknown length
            *   129456   143034: contig of 13579 bp in length
            *   143035   143134: gap of unknown length
            *   143135   160120: contig of 16986 bp in length
            *   160121   160220: gap of unknown length
            *   160221   169752: contig of 9532 bp in length
            *   169753   169852: gap of unknown length
            *   169853   177926: contig of 8074 bp in length
            *   177927   178026: gap of unknown length
            *   178027   186068: contig of 8042 bp in length
            *   186069   186168: gap of unknown length
            *   186169   192895: contig of 6727 bp in length
            *   192896   192995: gap of unknown length
            *   192996   198186: contig of 5191 bp in length
            *   198187   198286: gap of unknown length
            *   198287   203039: contig of 4753 bp in length
            *   203040   203139: gap of unknown length
            *   203140   205631: contig of 2492 bp in length
            *   205632   205731: gap of unknown length
            *   205732   206945: contig of 1214 bp in length
            *   206946   207045: gap of unknown length
            *   207046   209081: contig of 2036 bp in length
            *   209082   209181: gap of unknown length
            *   209182   210296: contig of 1115 bp in length
            *   210297   210396: gap of unknown length
            *   210397   212804: contig of 2408 bp in length
            *   212805   212904: gap of unknown length
            *   212905   214017: contig of 1113 bp in length
            *   214018   214117: gap of unknown length
            *   214118   216475: contig of 2358 bp in length
            *   216476   216575: gap of unknown length
            *   216576   217968: contig of 1393 bp in length
            *   217969   218068: gap of unknown length
            *   218069   219095: contig of 1027 bp in length
            *   219096   219195: gap of unknown length
            *   219196   220470: contig of 1275 bp in length
            *   220471   220570: gap of unknown length
            *   220571   221822: contig of 1252 bp in length
            *   221823   221922: gap of unknown length
            *   221923   223428: contig of 1506 bp in length
            *   223429   223528: gap of unknown length
            *   223529   224826: contig of 1298 bp in length
            *   224827   224926: gap of unknown length
            *   224927   226124: contig of 1198 bp in length
            *   226125   226224: gap of unknown length
            *   226225   227922: contig of 1698 bp in length
            *   227923   228022: gap of unknown length
            *   228023   229064: contig of 1042 bp in length
            *   229065   229164: gap of unknown length
            *   229165   230519: contig of 1355 bp in length
            *   230520   230619: gap of unknown length
            *   230620   231847: contig of 1228 bp in length
            *   231848   231947: gap of unknown length
            *   231948   233513: contig of 1566 bp in length
            *   233514   233613: gap of unknown length
            *   233614   234860: contig of 1247 bp in length
            *   234861   234960: gap of unknown length
            *   234961   235968: contig of 1008 bp in length
            *   235969   236068: gap of unknown length
            *   236069   237121: contig of 1053 bp in length
            *   237122   237221: gap of unknown length
```

```
FEATURES
    source
                                    * 237222  238257: contig of 1036 bp in length
                                    * 238258  238357: gap of unknown length
                                    * 238358  239413: contig of 1056 bp in length
                                    * 239414  239513: gap of unknown length
                                    * 239514  240850: contig of 1337 bp in length
                                    * 240851  240950: gap of unknown length
                                    * 240951  242140: contig of 1190 bp in length
                                    * 242141  242240: gap of unknown length
                                    * 242241  243504: contig of 1264 bp in length
                                    * 243505  243604: gap of unknown length
                                    * 243605  244749: contig of 1145 bp in length
                                    * 244750  244849: gap of unknown length
                                    * 244850  245880: contig of 1031 bp in length.
                   Location/Qualifiers
            1. .245880
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="12"
                   /clone="RP4-761J14, RP11-433J6"
BASE COUNT  61016 a  60422 c  59511 g  60793 t   4138 others
ORIGIN

Query Match    57.8%;   Score 288.8;  DB 2;  Length 245880;
Best Local Similarity  99.3%;   Pred. No. 2.3e-51;
Matches  290;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1 TCAGGCTGCCTGATCTGCCCAGCGTTTCCTCTGGATTCCGGCCCTCTGGTCATC  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123660 TCAGGCTGCCTGATCTGCCCAGCGTTTCCTCTGGATTCCGGCCCTCTGGTCATC 123601

Qy     61 CCTCCCCACCCTCTCTCTCCAAGGCGCCCTCTCTCTCGGTCTCCCCTTCTTCTTCTTAGAAACCCCTTCCTC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123600 CCTCCCCACCCTCTCTCTCCAAGGCGCCCTCTCTCTCGGTCTCCCCTTCTTCTTCTTAGAAACCCCTTCCTC 123541

Qy    121 CACCTCCCCTCTCTGCAGAACTTCTCCTTTACCCCCCACCCCACCACTGCCCCCTTTCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123540 CACCTCCCCTCTCTGCAGAACTTCTCCTTTACCCCCCACCCCACCACTGCCCCCTTTCC 123481

Qy    181 TTTTCTGACCTCCTTTTTGGAGGGCTCAGCGCTGCCCAGACCATAGGAGGAGTGTGGGAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123480 TTTTCTGACCTCCTTTTTGGAGGGCTCAGCGCTGCCCAGACCATAGGAGGAGTGTGGGAGG 123421

Qy    241 CTCAGTTCCTGGGGCTTGCTTGTTTTCTCCAGCCGCGCCTTTGGGTGCCTTCCAGTGAA 292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123420 CTCAGTTCCTGGGGCTTGCTTGTTTTCTCCAGCCGCGCCTTTGGGTGCCTTCCAGTAA 123369
```

Search completed: October 10, 2002, 18:30:58
Job time : 813.867 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2002, 12:39:37 ; Search time 715.867 Seconds
                 (without alignments)
                 14616.225 Million cell updates/sec

Title: US-09-489-101A-15_COPY_1_500
Perfect score: 500
Sequence: 1 gagtctacggcattgctgag.........cccagggcatcgcaacaag 500

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 gb_ba:*
                 gb_htg:*
                 gb_in:*
                 gb_om:*
                 gb_ov:*
                 gb_pat:*
                 gb_ph:*
                 gb_pl:*
                 gb_pr:*
                 gb_ro:*
                 gb_sts:*
                 gb_sy:*
                 gb_un:*
                 gb_vi:*
                 em_ba:*
                 em_fun:*
                 em_hum:*
                 em_in:*
                 em_mu:*
                 em_om:*
                 em_or:*
                 em_ov:*
                 em_pat:*
                 em_ph:*
                 em_pl:*
                 em_ro:*
                 em_sts:*
                 em_un:*
                 em_vi:*
                 em_htg_hum:*
                 em_htg_inv:*
                 em_htg_other:*
                 em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result   Score   Query   Length DB   ID   Description
NO.   Match

No.	c	Score	%	Length	n	DB ID	Description
1		500	100.0	748	6	AX201595	AX201595 Sequence
2		500	100.0	1201	6	AX201594	AX201594 Sequence
3		498.4	99.7	2448	9	HSM801408	AL133561 Homo sapi
4		359.4	71.9	157086	9	AL355885	AL355885 Homo sapi
5		357.8	71.6	150754	9	AC023491	AC023491 Homo sapi
6		356.2	71.2	150724	9	AP000552	AP000552 Homo sapi
7	c	354.6	70.9	122364	9	AP000550	AP000550 Homo sapi
8	c	349.8	70.0	159550	2	AC023491	AC023491 Homo sapi
9	c	349.8	70.0	159550	2	AC007324	AC007324 Homo sapi
10	c	349.6	70.0	98557	2	AC013360	AC013360 Homo sapi
11	c	348.2	69.6	106650	9	AC013360	AC013360 Homo sapi
12	c	346.6	69.3	184590	2	AC008103	AC008103 Homo sapi
13	c	345	69.0	180876	9	AC007708	AC007708 Homo sapi
14	c	344	68.8	175358	9	AC009288	AC009288 Homo sapi
15	c	325.4	65.1	180884	9	AC007981	AC007981 Homo sapi
16	c	324.8	65.0	182617	9	AC007981	AC007981 Homo sapi
17	c	324.8	65.0	192592	9	AC008018	AC008018 Homo sapi
18	c	321	64.2	38849	9	AC007731	AC007731 Homo sapi
19	c	310	62.0	165050	9	AC005500	AC005500 Homo sapi
20	c	307.8	61.6	170102	9	AC024070	AC024070 Homo sapi
21	c	297.8	59.6	1647	9	AC007325	AC007325 Homo sapi
22	c	297.8	59.6	1647	9	AC008079	AC008079 Homo sapi
23	c	265	53.0	28862	9	HSM800994	HSM800994 Homo sapi
24	c	265	53.0	176051	9	AL117481	AL117481 Homo sapi
25	c	259.6	51.9	145598	9	AL122069	AL122069 Homo sapi
26	c	250	50.0	94384	9	AL122069	AL122069 Homo sapi
27	c	229.6	45.9	175358	9	AC012398	AC012398 Homo sapi
28	c	180.2	36.0	202450	9	AC023490	AC023490 Homo sapi
29	c	177.6	35.5	160866	9	AC008132	AC008132 Homo sapi
30	c	171.6	34.3	69648	2	AC011718	AC011718 Homo sapi
31	c	131.2	26.2	110000	2	AC007981	AC007981 Homo sapi
32	c	116.4	23.3	390	6	AC090966	AC090966 Homo sapi
33	c	112.8	22.6	749	14	AC092532	AC092532 Papio cyn
34	c	112.8	22.6	131995	1	AC090966	AC090966 Papio cyn
35	c	93.6	18.7	47852	1	AC092532	AC092532 Papio cyn
36	c	92.8	18.6	184590	2	AC099825	AC099825 Papio ham
37	c	91.4	18.3	112509	2	AR135147	AR135147 Sequence
38	c	90	18.0	1647	9	AL353694_1	AL353694_1 Continuation (2 of
39	c	90	18.0	1647	9	NPORP2	NPORP2
40	c	90	18.0	1647	9	OPU75930	OPU75930 Orgyia pseu
41	c	88.6	17.7	15348	1	U75930	U75930 Orgyia pseu
42	c	88	17.6	2448	1	D13755	D13755 Multicapsid
43	c	87.4	17.5	15311	1	AL022022	AL022022 Mycobacte
44	c	87.4	17.5	47852	1	MTV023	AB042821 Mycobacte
45	c	87.2	17.4	1950	1	AB042821	AB042821 Streptomy
	c	87	17.4	122364	9	AC007324	AC007324 Homo sapi

ALIGNMENTS

RESULT 1
AX201595
AX201595

LOCUS   AX201595   748 bp   DNA   linear   PAT 30-AUG-2001
DEFINITION Sequence 16 from Patent WO0153349.
ACCESSION   AX201595
VERSION   AX201595.1 GI:15391445
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 748)
  AUTHORS   Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
            Chen,Y.T.
  TITLE   Small cell lung cancer associated antigens and uses therefor
  JOURNAL   Patent: WO 0153349-A 16 26-JUL-2001;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US) ; MEMORIAL
            SLOAN-KETTERING CANCER CENTER (US) ; CORNELL RESEARCH FOUNDATION,
            INC. (US)
FEATURES   Location/Qualifiers
     source   1..748
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

```
BASE COUNT      166 a      261 c      246 g       72 t        3 others
ORIGIN

Query Match     100.0%;   Score 500;  DB 6;  Length 748;
Best Local Similarity   100.0%;  Pred. No. 1.8e-36;
Matches   500;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGAGGACGCCGACCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGAGGACGCCGACCAGG 60

Qy     61 GCATCGCTAATGAGGAGGACACCACCCAGTGCATCGCTAATGAGGAGGACGCCGACCATCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATCGCTAATGAGGAGGACACCACCCAGTGCATCGCTAATGAGGAGGACGCCGACCATCG 120

Qy    121 CCGAGGACGCCATCCAGGGGCATCGCCAAACGAGGAGTTGCCCAGGGCATCGCCAAATGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGGACGCCATCCAGGGGCATCGCCAAACGAGGAGTTGCCCAGGGCATCGCCAAATGGGG 180

Qy    181 TCGCGCCACAGGGGCATCGCCAAATGAGGACGCGCCCAACGAGGAGCGCCCAACTGGGACGCCG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGCGCCACAGGGGCATCGCCAAATGAGGACGCGCCCAACGAGGAGCGCCCAACTGGGACGCCG 240

Qy    241 TCCACGGCTTCGCCAACGAGGTCACCATCCACCACGCCATCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCACGGCTTCGCCAACGAGGTCACCATCCACCACGCCATCG 300

Qy    301 GCATCGCCAACGGGGGACGCCATGGGCGCAACGAGGTCAACGAGCGCCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCATCGCCAACGGGGGACGCCATGGGCGCAACGAGGTCAACGAGCGCCATCG 360

Qy    361 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCAGGGCATCGCCAACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCAGGGCATCGCCAACG 420

Qy    421 AGGACGCCGCGCCCAGGGAATCGCCGAGGAGTGTCGCACAGGGGCATCGCCAACGAGGACGCCG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGACGCCGCGCCCAGGGAATCGCCGAGGAGTGTCGCACAGGGGCATCGCCAACGAGGACGCCG 480

Qy    481 CCCAGGGGCATCGCCAACAAG 500
          |||||||||||||||||||||
Db    481 CCCAGGGGCATCGCCAACAAG 500
```

```
RESULT 2
LOCUS           AX201594              1201 bp   DNA      linear     PAT 30-AUG-2001
DEFINITION      Sequence 15 from Patent WO0153349.
ACCESSION       AX201594
VERSION         AX201594.1   GI:15914413
KEYWORDS        .
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 1201)
  AUTHORS       Stockert,E., Scanlan,M.J., Jager,D., Old,L.J., Gure,A.O. and
                Chen,Y.T.
  TITLE         Small cell lung cancer associated antigens and uses therefor
  JOURNAL       Patent: WO 0153349-A 15 26-JUL-2001;
                LUDWIG INSTITUTE FOR CANCER RESEARCH (US); MEMORIAL
                SLOAN-KETTERING CANCER CENTER (US); CORNELL RESEARCH FOUNDATION,
                INC. (US)
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      255 a      377 c      349 g      178 t       42 others
ORIGIN

Query Match     100.0%;   Score 500;  DB 6;  Length 1201;
Best Local Similarity   100.0%;  Pred. No. 1.6e-36;
Matches   500;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy      1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGAGGACGCCGACCAGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGAGGACGCCGACCAGG 60

Qy     61 GCATCGCTAATGAGGAGGACACCACCCAGTGCATCGCTAATGAGGAGGACGCCGACCATCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATCGCTAATGAGGAGGACACCACCCAGTGCATCGCTAATGAGGAGGACGCCGACCATCG 120

Qy    121 CCGAGGACGCCATCCAGGGGCATCGCCAAACGAGGAGTTGCCCAGGGCATCGCCAAATGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGGACGCCATCCAGGGGCATCGCCAAACGAGGAGTTGCCCAGGGCATCGCCAAATGGGG 180

Qy    181 TCGCGCCACAGGGGCATCGCCAAATGAGGACGCGCCCAACGAGGAGCGCCCAACTGGGACGCCG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGCGCCACAGGGGCATCGCCAAATGAGGACGCGCCCAACGAGGAGCGCCCAACTGGGACGCCG 240

Qy    241 TCCACGGCTTCGCCAACGAGGTCACCATCCACCACGCCATCG 300
          ||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCACGGCTTCGCCAACGAGGTCACCATCCACCACGCCATCG 300

Qy    301 GCATCGCCAACGGGGGACGCCATGGGCGCAACGAGGTCAACGAGCGCCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCATCGCCAACGGGGGACGCCATGGGCGCAACGAGGTCAACGAGCGCCATCG 360

Qy    361 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCAGGGCATCGCCAACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCAGGGCATCGCCAACG 420

Qy    421 AGGACGCCGCGCCCAGGGAATCGCCGAGGAGTGTCGCACAGGGGCATCGCCAACGAGGACGCCG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGACGCCGCGCCCAGGGAATCGCCGAGGAGTGTCGCACAGGGGCATCGCCAACGAGGACGCCG 480

Qy    481 CCCAGGGGCATCGCCAACAAG 500
          |||||||||||||||||||||
Db    481 CCCAGGGGCATCGCCAACAAG 500
```

```
RESULT 3
LOCUS           HSM801408             2448 bp   mRNA     linear     PRI 18-FEB-2000
DEFINITION      Homo sapiens mRNA; cDNA DKFZp434C196 (from clone DKFZp434C196);
                partial cds.
ACCESSION       AL133561
VERSION         AL133561.1   GI:6599133
KEYWORDS        .
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 2448)
  AUTHORS       Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE         Direct Submission
  JOURNAL       Submitted (15-DEC-1999) MIPS, Am Klopferspitz 18a, D-82152
                Martinsried, GERMANY
COMMENT         Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
                Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
                sequenced by BMFZ (Biomedical Research Center at the Charite,
                Berlin/Germany) within the cDNA sequencing consortium of the German
                Genome Project.
                This clone (DKFZp434C196) is available at the RZPD in Berlin.
                Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
                Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
                information about the clone and the sequencing project is available
                at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..2448
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434C196"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3)"
                     /clone_lib="DKFZp434C196"
                     DH10B; sites NotI + SalI"
```

```
Qy      1 GCATCGCTAATGAGGACACCACCCAGTGCATCGCTAATGAGGAGGACGCCGACCATCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGTCTACGGCATTGCTGAGGACGCTGCCCAGGGCATCGCTAATGAGGAGGACGCCGACCAGG 60

Qy     61 CCGAGGACGCCATCCAGGGGCATCGCCAAACGAGGAGTTGCCCAGGGCATCGCCAAATGGGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATCGCTAATGAGGAGGACACCACCCAGTGCATCGCTAATGAGGAGGACGCCGACCATCG 120

Qy    121 TCGCGCCACAGGGGCATCGCCAAATGAGGACGCGCCCAACGAGGAGCGCCCAACTGGGACGCCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGAGGACGCCATCCAGGGGCATCGCCAAACGAGGAGTTGCCCAGGGCATCGCCAAATGGGG 180

Qy    181 TCCACGGCTTCGCCAACGAGGTCACCATCCACCACGCCATCG 240
          ||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGCGCCACAGGGGCATCGCCAAATGAGGACGCGCCCAACGAGGAGCGCCCAACTGGGACGCCG 240

Qy    241 GCATCGCCAACGGGGGACGCCATGGGCGCAACGAGGTCAACGAGCGCCATCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCACGGCTTCGCCAACGAGGTCACCATCCACCACGCCATCG 300

Qy    301 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCAGGGCATCGCCAACG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCATCGCCAACGGGGGACGCCATGGGCGCAACGAGGTCAACGAGCGCCATCG 360

Qy    361 AGGACGCCGCGCCCAGGGAATCGCCGAGGAGTGTCGCACAGGGGCATCGCCAACGAGGACGCCG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTAACGAGGACGCCGTCCAGGGCATCGCTAACGAGGTGGCCCAGGGCATCGCCAACG 420

Qy    421 AGGACGCCGCGCCCAGGGAATCGCCGAGGATGTGCGCACAGGGGCATCGCCAACGAGGACGCCG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGACGCCGCGCCCAGGGAATCGCCGAGGAGTGTCGCACAGGGGCATCGCCAACGAGGACGCCG 480

Qy    481 CCCAGGGGCATCGCCAACAAG 500
          |||||||||||||||||||||
Db    481 CCCAGGGGCATCGCCAACAAG 500
```

```
ACCESSION   AL356585
DEFINITION  Homo sapiens chromosome 13 clone RP11-341D18, *** SEQUENCING IN
            PROGRESS ***, 13 unordered pieces.
LOCUS       AL356585          184590 bp    DNA     linear   HTG 20-JAN-2001
RESULT 4
AL356585/c
```

```
gene            /dev_stage="adult"
                1..1744
                /gene="DKFZp434C196"
CDS             <1..1744
                /gene="DKFZp434C196"
                /note="similar to mucins"
                /codon_start=2
                /product="hypothetical protein"
                /protein_id="CAB63715.1"
                /db_xref="GI:6599134"
                /translation="SPSRASLTRTPPRASLMRPSPASLTRTPSRASPTRMPSRASLK
                MTPRASLTRKMESTALLRTLPRASLMRTPTRASLMRTPPRASPTRKPPRASPTPSRA
                SPTRRLPRASPMGSPHRASPMRTPPRASPTGTPSTSASPTGTPSSASPTGTPPRASPTG
                TPPRAWATRSPSTASLTRTPSRASLTRWPPRASPTRTPPRESPRMSHRASPTRTPPRA
                SPTRRPPWATRSPSTASLTRTSHRASPTRMPPRASPTRRPPRASPTGSPPRASPMTP
                PRASPRPPWASPTPPRASPTTTPRASLTRTPSWASPTTTPSRASLMKMESTVSITRTPPRASP
                TGTPSRASPGTPSRASLTGSPSRASLIGTPSRASLTGTPSRASLIGTPSRASLTGTP
                RASLTGTSSTASLTRTPQSSSSLTRTPSMASLTRTPPPRASLITRTPPRAS
                LTRTPPRASLITRTPPRASLMTPSMVSLKRSPSRASLTRTPSRASLITRTPSRASLTRT
                PSTASLTGTPPTASLITRTPPTASLTRTPSPPTASLTRTPSTASLTRMPSTASLTRKSNVN
                QQCPASTPSSEVIS"
polyA_signal    2397..2402
polyA_site      2418
ORIGIN
BASE COUNT    643 a    727 c    648 g    430 t

Query Match        99.7%;   Score 498.4;   DB 9;   Length 2448;
Best Local Similarity   99.8%;   
Matches 499;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy       1  GAGTCTACGGCATTGCTGAGACGCGCTCCCCAGGCGCATCGCTAATGAGGACGCCGACCAGG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     167  GAGTCTACGGCATTGCTGAGACGCGCTCCCCAGGCGCATCGCTAATGAGGACGCCGACCAGG  226

Qy      61  GCATCGCTAATGAGGACACCACCCAGTGCATCGCTCGCCAGGCAGGAAGCGCCCAGGCATCG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     227  GCATCGCTAATGAGGACACCACCCAGTGCATCGCTCGCCAGGCAGGAAGCGCCCAGGCATCG  286

Qy     121  CCGAGGACGCGCCATCCAGGGCATCGCCAACGAGACGAGAGGTTGCCCAGGCCATCGCCAATGGGG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     287  CCGAGGACGCGCCATCCAGGGCATCGCCAACGAGACGAGAGGTTGCCCAGGCCATCGCCAATGGGG  346

Qy     181  TCGCCGCCCACCAGGGCATCGCCAATGAGGACGCGCCACCACCCAGCGCCACCACTGGGACGCCG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     347  TCGCCGCCCACCAGGGCATCGCCAATGAGGACGCGCCACCACCCAGCGCCACCACTGGGACGCCG  406

Qy     241  TCCACGGCGTTCGCCAACGGGGACGCCGTCCTCAGCTTCGCCAACGGGGACGCCGCCCAGG  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     407  TCCACGGCGTTCGCCAACGGGGACGCCGTCCTCAGCTTCGCCAACGGGGACGCCGCCCAGG  466

Qy     301  GCATCGCCAACGGGGACGCCACCAAGGGCATGGGCGAACGAGGTCACCATCCACGGCATCG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     467  GCATCGCCAACGGGGACGCCACCAAGGGCATGGGCGAACGAGGTCACCATCCACGGCATCG  526

Qy     361  CTAACGAGGAGGACGCGCCGTCAGGGCATCGCTAACGAGGTGGGCGCGCCCCAGGGCAACG  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527  CTAACGAGGAGGACGCGCCGTCAGGGCATCGCTAACGAGGTGGGCGCGCCCCAGGGCAACG  586

Qy     421  AGGACGCCGCCCAGGGAATCGCCGAGGAGTCGCACGAGGACGCGGCG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db     587  AGGACGCGCGCCCAGGGAATCGCCGAGGAGTCGCACGAGGACGCGGCG  646

Qy     481  CCCAGGGCATCGCCAACAAG  500
            ||||||||||||||||||||
Db     647  CCCAGGGCATCGCCAACAAG  666
```

```
VERSION     AL355585.3  GI:9800966
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184590)
AUTHORS     Burton,J.
TITLE       Direct Submission
JOURNAL     Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Aug 14, 2000 this sequence version replaced gi:9213941.
COMMENT     -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA341D18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing program: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 179028 bases at least Q40
            Consensus quality: 181067 bases at least Q30
            Consensus quality: 182219 bases at least Q20
            Insert size: 183390; sum-of-contigs
            Insert size: 192611; 2.8% error; agarose-fp
            Quality coverage: 4.33x in Q20 bases; sum-of-contigs Quality
            coverage: 4.20x in Q20 bases; agarose-fp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*          1   36485: contig of 36485 bp in length
*      36486   36585: gap of 100 bp
*      36586   61226: contig of 2461 bp in length
*      61227   61326: gap of 100 bp
*      61327   86064: contig of 24738 bp in length
*      86065   86164: gap of 100 bp
*      86165   92534: contig of 6370 bp in length
*      92535   92634: gap of 100 bp
*      92635   97289: contig of 4655 bp in length
*      97290   97389: gap of 100 bp
*      97390  100517: contig of 3128 bp in length
*     100518  100617: gap of 100 bp
*     100618  116404: contig of 15787 bp in length
*     116405  116504: gap of 100 bp
*     116505  124935: contig of 8431 bp in length
*     124936  125035: gap of 100 bp
*     125036  127757: contig of 2722 bp in length
*     127758  127857: gap of 100 bp
*     127858  140183: contig of 12326 bp in length
*     140184  140283: gap of 100 bp
*     140284  143272: contig of 2989 bp in length
*     143273  143372: gap of 100 bp
*     143373  156929: contig of 13557 bp in length
*     156930  157029: gap of 100 bp
*     157030  184590: contig of 27561 bp in length.
FEATURES             Location/Qualifiers
     source          1..184590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-341D18"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..36485
```

```
misc_feature      /note="assembly_fragment:00721"
                  fragment_chain:1
                  clone_end:SP6
                  vector_side:left

misc_feature      36586..61226
                  /note="assembly_fragment:01482"
                  fragment_chain:1

misc_feature      61327..86064
                  /note="assembly_fragment:00458"
                  fragment_chain:1

misc_feature      86165..92534
                  /note="assembly_fragment:00373"
                  fragment_chain:1

misc_feature      92635..97289
                  /note="assembly_fragment:00679"
                  fragment_chain:1

misc_feature      97390..100517
                  /note="assembly_fragment:01224"
                  fragment_chain:2

misc_feature      100618..116404
                  /note="assembly_fragment:00877"
                  fragment_chain:2

misc_feature      116505..124935
                  /note="assembly_fragment:00503"
                  fragment_chain:2

misc_feature      125036..127757
                  /note="assembly_fragment:00226"
                  fragment_chain:2

misc_feature      127858..140183
                  /note="assembly_fragment:01286"
                  fragment_chain:2

misc_feature      140284..143272
                  /note="assembly_fragment:01214"
                  fragment_chain:2

misc_feature      143373..156929
                  /note="assembly_fragment:01825"
                  fragment_chain:2

misc_feature      157030..184590
                  /note="assembly_fragment:00863"
                  clone_end:T7
                  vector_side:right

BASE COUNT    49855 a   39974 c   41941 g   51612 t   1208 others
ORIGIN

Query Match  99.7%;  Best Local Similarity  99.8%;
Score 498.4;  DB 2;  Length 184590;
Matches 499;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;
Pred. No. 4.6e-37;

Qy        1 GAGTCTACGGACGTTGCTGAGGACGGCTGCCCAGGGGCATCGTCCTAATGAGACGACGCCAGG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123163 GAGTCTACGGACGTTGCTGAGGACGGCTGCCCAGGGGCATCGTCCTAATGAGACGACGCCAGG 123104

Qy       61 GCATCCGCTTAATGAGGACACACCCAGTGCCATCGCCGCCAGGGAAGCCGCCGCCACCAATCG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123103 GCATCCGCTTAATGAGGACACACCCAGTGCCATCGCCGCCAGGGAAGCCGCCGCCACCAATCG 123044

Qy      121 CCGAGGAGACGCCATCCAGGGCATCGCCAACGAGGAGTTGCCCAGGGCATCGCCAATGGGG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123043 CCGAGGAGACGCCATCCAGGGCATCGCCAACGAGGAGTTGCCCAGGGCATCGCCAATGGGG 122984

Qy      181 TCGCCCGCACAGGGCATCGCCAATGAGGACGACGCCACCACGAGGACTGGGACGCCG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122983 TCGCCCGCACAGGGCATCGCCAATGAGGACGACGCCACCACGAGGACTGGGACGCCG 122924

Qy      241 TCCACGCGCTTCGCCAACGGGGACGCCGCCTCCTCAGCGTTCGCCAACGGGGACGCCGCCAGG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122923 TCCACGCGCTTCGCCAACGGGGACGCCGCCTCCTCAGCGTTCGCCAACGGGGACGCCGCCAGG 122864

Qy      301 GCATCGCCCACAGGGGACGGGGACGCCCACCAAGGGCATGGGCAAGGAGGTCACCATCCACGGCATCG 360
            |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   122863 GCATCGCCCACAGGGGACGCACCAAGGGCATGGGCAAGGAGGTCACCATCCACGGCATCG 122804

Qy      361 CTAACGAGGAGGACCGTCCAGGGCATCGCTAACGAGGAGGTGGCCGCGCCCAGGGGCATCGCCAACG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122803 CTAACGAGGAGGACCGTCCAGGGCATCGCTAACGAGGAGGTGGCCGCGCCCAGGGGCATCGCCAACG 122744

Qy      421 AGGACGACGCCGCCGAGAATCGCCGAGGATGTCGCCACACGGGCATCGCCAACGAGGACGCCG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122743 AGGACGACGCCGCCGAGAATCGCCGAGGATGTCGCCACACGGGCATCGCCAACGAGGACGCCG 122684

Qy      481 CCCAGGGCATCGCCAACAAG 500
            ||||||||||||||||||||
Db   122683 CCCAGGGCATCGCCAACAAG 122664
```

RESULT 5

```
LOCUS       AP000552/c    157086 bp    DNA    linear    PRI 01-OCT-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
            clone:KB1183D5.
ACCESSION   AP000552
VERSION     AP000552.1  GI:5931538
KEYWORDS
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:Keio
            BAC library clone:KB1183D5.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157086)
  AUTHORS   Shimizu,N.
  TITLE     Human DNA sequence from clone KB1183D5 on chromosome 22q11.2
  JOURNAL   Published Only in Database (1999) In press
REFERENCE   2  (bases 1 to 157086)
  AUTHORS   Shimizu,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1999) to the DDBJ/EMBL/GenBank databases.
            Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
            Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan
            (E-mail:nshinizu@dmb.med.keio.ac.jp, Tel:81-3-3351-2370,
            Fax:81-3-3351-2370)
COMMENT     This is a complete sequence of the clone. The
            proximal adjacent clone is KB876E2 (Acc. #AP000551) with 379-bp
            overlapping.
FEATURES             Location/Qualifiers
     source          1..157086
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="FLEB14-14"
                     /cell_type="pre-pro-B cell"
                     /chromosome="22"
                     /clone="KB1183D5"
                     /clone_lib="Keio BAC library"
                     /map="22q11.2"
     repeat_region   57..355
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   complement(1494..1585)
                     /rpt_family="MLT1iI"
                     /evidence=not_experimental
     repeat_region   complement(2609..2680)
                     /rpt_family="MLT1L"
                     /evidence=not_experimental
     repeat_region   complement(3880..4025)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   9616..9787
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   9689..9790
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   9794..9819
                     /rpt_family="(A)n"
                     /evidence=not_experimental
     repeat_region   complement(10704..10984)
                     /rpt_family="AluY"
```

```
repeat_region   /evidence=not_experimental
                11054. .11343
repeat_region   /rpt_family="AluSx"
                /evidence=not_experimental
                11344. .11367
repeat_region   /rpt_family="(TTAA)n"
                /evidence=not_experimental
                complement(12418. .12736)
repeat_region   /rpt_family="AluSq"
                /evidence=not_experimental
                complement(13993. .14328)
repeat_region   /rpt_family="AluJo"
                /evidence=not_experimental
                complement(14759. .14875)
repeat_region   /rpt_family="FLAM_C"
                /evidence=not_experimental
                14876. .14886
repeat_region   /rpt_family="Charlie4"
                /evidence=not_experimental
                complement(14887. .15155)
repeat_region   /rpt_family="AluSx"
                /evidence=not_experimental
                15156. .15239
repeat_region   /rpt_family="Charlie4"
                /evidence=not_experimental
                15618. .15788
repeat_region   /rpt_family="L2"
                /evidence=not_experimental
                15911. .15924
repeat_region   /rpt_family="L1MB7"
                /evidence=not_experimental
                complement(15925. .16230)
repeat_region   /rpt_family="AluJo"
                /evidence=not_experimental
                16231. .16940
repeat_region   /rpt_family="L1MB7"
                /evidence=not_experimental
                complement(17696. .17816)
repeat_region   /rpt_family="FLAM_A"
                /evidence=not_experimental
                complement(17861. .18018)
repeat_region   /rpt_family="AluJo"
                /evidence=not_experimental
                18019. .18049
repeat_region   /rpt_family="(TTTTA)n"
                /evidence=not_experimental
                complement(18050. .18176)
repeat_region   /rpt_family="AluJo"
                /evidence=not_experimental
                18300. .18602
repeat_region   /rpt_family="AluJo"
                /evidence=not_experimental
                complement(19466. .19612)
repeat_region   /rpt_family="L2"
                /evidence=not_experimental
                19721. .20022
repeat_region   /rpt_family="Aluy"
                /evidence=not_experimental
                complement(20025. .20302)
repeat_region   /rpt_family="AluSg"
                /evidence=not_experimental
                complement(20308. .20325)
repeat_region   /rpt_family="AluSp/q"
                /evidence=not_experimental
                20326. .20374
repeat_region   /rpt_family="(TG)n"
                /evidence=not_experimental
                complement(20375. .20445)
repeat_region   /rpt_family="AluSp/q"
                /evidence=not_experimental
                20593. .20669
repeat_region   /rpt_family="AluSp/q"
                /evidence=not_experimental
```

```
repeat_region   complement(20885. .21149)
                /rpt_family="AluY"
                /evidence=not_experimental
repeat_region   21186. .21481
                /rpt_family="AluY"
                /evidence=not_experimental
repeat_region   complement(21482. .21758)
                /rpt_family="AluSg"
                /evidence=not_experimental
repeat_region   complement(21759. .21862)
                /rpt_family="AluSp/q"
                /evidence=not_experimental
repeat_region   24439. .24588
                /rpt_family="AluY"
                /evidence=not_experimental
repeat_region   complement(24678. .24968)
                /rpt_family="AluSx"
                /evidence=not_experimental
repeat_region   complement(25016. .25306)
                /rpt_family="AluYb8"
                /evidence=not_experimental
repeat_region   complement(25310. .25609)
                /rpt_family="AluYa8"
                /evidence=not_experimental
repeat_region   25627. .25853
                /rpt_family="AluSx"
                /evidence=not_experimental
repeat_region   25858. .25885
                /rpt_family="(A)n"
                /evidence=not_experimental
repeat_region   26308. .26588
                /rpt_family="AluJo"
                /evidence=not_experimental
repeat_region   complement(30748. .31041)
                /rpt_family="AluJo"
                /evidence=not_experimental
repeat_region   35559. .35661
                /rpt_family="AluSg"
                /evidence=not_experimental
repeat_region   complement(35833. .35996)
                /rpt_family="AluSg/x"
                /evidence=not_experimental
repeat_region   36421. .36721
                /rpt_family="MER5B"
                /evidence=not_experimental
repeat_region   37016. .37083
                /rpt_family="AluSg"
                /evidence=not_experimental
repeat_region   37175. .37372
                /rpt_family="MLT1E"
                /evidence=not_experimental
repeat_region   37480. .37793
                /rpt_family="MLT1E2"
                /evidence=not_experimental
repeat_region   38069. .38351
                /rpt_family="AluY"
                /evidence=not_experimental
repeat_region   38506. .38612
                /rpt_family="AluSg"
                /evidence=not_experimental
repeat_region   38841. .38985
                /rpt_family="MIR"
                /evidence=not_experimental
repeat_region   38987. .39118
                /rpt_family="AluSg/x"
                /evidence=not_experimental
repeat_region   39119. .39173
                /rpt_family="L1PA16"
                /evidence=not_experimental
                /rpt_family="L1PA16"
```

Query Match   71.9%;   Score 359.4; DB 9;   Length 157086;
Best Local Similarity 86.3%;   Pred. No. 1.1e-24;
Matches   422;   Conservative   0;   Mismatches   61;   Indels   6;   Gaps   2;

```
Qy      18  GAGGACGCGTGCCCAGGGCATCCGTTAATGAGGACGGGGGCATCCGTTAATGAGGAC     77
            |||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   48981  GAGGACGCGCCCCAGGGCATCCGTTAATGAGGACGCCGCCAACGAGGAC          48922

Qy      78  ACCACCCAGTGCATCGCCAACGAGGAAGCCCCAGGGCATCGCC---GAGGACGCCATC    134
            ||  |||| |||  |||||||||||||||||||||||||||||   ||||||||||||
Db   48921  GCCGCCCAGGGCATCGCCAACGAGGACGACGCCCAGGGCATCGCC--GAGGACGCCGCC 48862

Qy     135  CAGGGCCATCCCAACGAGGAGGTTGCCCAGGGCATCGCCGGGTCGCGCGCACAGGGC    194
            ||||||||||||||||||||| | || |||||||| |||||||||||||||||||||
Db   48861  CAGGGCCATCCCAACGAGGACGCGCCCCAGGGCATCGCCAAAGAGGAC           48802

Qy     195  ATCGCCAATGAGGACGCACCCAGGGCCATCCCAACGAGGCGCCGTTCGCC          254
            |||||| |||||||||||||||||||| ||||||||||||||||||||||
Db   48801  ATCGCCAAGGAGGACGCGCCCCAGGGCCATCCCAACGAGGACGCCGCC          48742

Qy     255  AACGGGGGACGCCGTCCTCAGCTTCGCCCAGGGCATCGCCCAGGGCATCGCCCAGGGC    314
            ||  ||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db   48741  AAAGAGGACGCCGTCCCAGGGCCATCGCCCAACGAGGACGCCCAGGGCATCGCCAACGAG 48682

Qy     315  GACGCCACCAAGGGCATGGGCAACGAGGTCACCATCCCACGGCGTTAACGAGGACGCC    374
            |||||||||||||| |||||||||||| ||| ||||||||||| ||||||||||||||
Db   48681  GACGCCGCGCCACCGGAATCGCCCACGGAGGCCGCGCGAGGACGCC               48622

Qy     375  GTCCAGGGGCCATCGCTAAGCACGGTGGCGCCCCCAGGGCATCGCCGAGGACGCGCCC    434
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48621  GTCCAGGGCATCGCCAAGAGGATGCCGCGCCAGGGCATCGCCCAGGGCATCGCCCCCAG 48562

Qy     435  GGAATCGC---CGAGGATGTGCGCACAGGGAGCATCGCCCAGGGCATC             491
            ||| |||    |||| || |||||||||||||||||||||||||||||
Db   48561  GGCATCGCCAACGAGGACGCGCCGCCGCGGGCATCGCCCAGGGCATC           48502

Qy     492  GCCAACAAG    500
            |||||||
Db   48501  GCCAACGAG 48493
```

RESULT 6

Query Match   71.6%; Score 357.8; DB 9; Length 150754;
Best Local Similarity 86.1%;
Matches 421; Conservative 0; Mismatches 62; Indels 6; Gaps 2;
Pred. No. 1.5e-24;

```
Qy      18  GAGGACGCGTGCCCAGGGCATCGCTAATGAGGACGGGGGCATCGCTTCCACGGCTTCGCC    77
Db  111053  ATCGCCAACGAGGACGCGCCCCAGGGCATCGCCAACTGGGACGGCGCCTTCCACGGCGCC 110994

Qy      78  ACCACCCAGTGCATCGCCAACGAGGAAGCCCCAGGGCATCGCC---GAGGACGCCATC   134
Db  111233  GAGGACGCGCCCCAGGGCATCGCCAACGAGGACGACGCCCAGGGCATCGCCAAAGAGGAC 111174

Qy     135  CAGGGCCATCGCCAACGAGGAGGTTGCCCAGGGCATCGCCAATGGGGTGCGCGCACAGGGC   194
Db  111173  GCCGCCCCAGGGCATCGCCAACGAGGAGCCCCAGGGCATCGCCAAAGAGGACGCGCCC 111114

Qy     195  ATCGCCAATGAGGACGCACCCAGGGCCATCGCCAACGAGGCGCCGTTCGCC          254
Db  111053  ATCGCCAACGAGGACGCGCCCCAGGGCATCGCCAACGAGGACGCCGCC          110994

Qy     255  AACGGGGGACGCCGTCCTCAGCTTCGCCCAGGGCATCGCCCAGGGCATCGCC          314
Db  110933  GACGCCGCGCCACCGGCATCGCCAACGAGGACGCCCAGGGCATCGCC          110874

Qy     315  GACGCCACCAAGGGCATGGGCAACGAGGTCACCATCCCACGGCGTTAACGAGGACGCC    374
Db  110933  GACGCCGCGCCACCGGCATCGCCCACGGAGGCCGCGCGAGGACGCC          110874

Qy     375  GTCCAGGGGCCATCGCTAAGCACGGTGGCGCCCCCAGGGCATCGCCGAGGACGCGCCC    434
Db  110873  GTCCACGGCATCGCCAAGAGGATGCCGCGCCAGGGCATCGCCCAGGGCATCGCCCCAG 110814

Qy     435  GGAATCGC---CGAGGATGTCGCACAGGGAGCATCGCCCAGGGCATC             491
Db  110813  GGCATCGCCAACGAGGACGCGCCGCCGCGGGCATCGCCCAGGGCATC           110754

Qy     492  GCCAACAAG    500
Db  110753  GCCAACGAG 110745
```

RESULT 7

```
REFERENCE   1  (bases 1 to 150724)
AUTHORS     Shimizu,N.
TITLE       Human DNA sequence from clone KB1592A4 on chromosome 22q11.2
JOURNAL     Published Only in Database (1999) In press
REFERENCE   2  (bases 1 to 150724)
AUTHORS     Shimizu,N.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1999) to the DDBJ/EMBL/GenBank databases.
            Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
            Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan
            (E-mail:nshimizu@dmb.med.keio.ac.jp, Tel:81-3-3351-2370,
            Fax:81-3-3351-2370)
COMMENT     This is a complete sequence of the insert of KB1592A4 clone. The
            distal adjacent clone is KB876E2 (Acc.#AP000551) with 1196-bp
            overlapping.
FEATURES             Location/Qualifiers
source          1..150724
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="FLEB14-14"
                /cell_type="pre-pro-B cell"
                /chromosome="22"
                /clone="KB1592A4"
                /clone_lib="Keio BAC library"
                /map="22q11.2"
repeat_region   complement(1..90)
                /rpt_family="AluS"
                /evidence=not_experimental
repeat_region   complement(99..399)
                /rpt_family="AluSg"
                /evidence=not_experimental
repeat_region   complement(401..570)
                /rpt_family="LTR33"
                /evidence=not_experimental
repeat_region   complement(572..640)
                /rpt_family="MIR"
                /evidence=not_experimental
repeat_region   649..716
                /rpt_family="(TCCA)n"
                /evidence=not_experimental
repeat_region   complement(717..752)
                /rpt_family="L2"
                /evidence=not_experimental
repeat_region   complement(757..862)
                /rpt_family="MIR"
                /evidence=not_experimental
repeat_region   complement(933..1059)
                /rpt_family="L2"
                /evidence=not_experimental
repeat_region   1955..2151
                /rpt_family="AluJb"
                /evidence=not_experimental
repeat_region   2578..2875
                /rpt_family="AluSx"
                /evidence=not_experimental
repeat_region   complement(3310..3408)
                /rpt_family="LTR16A"
                /evidence=not_experimental
repeat_region   4424..4591
                /rpt_family="L1PA4"
                /evidence=not_experimental
repeat_region   4617..4637
                /rpt_family="(TAAA)n"
                /evidence=not_experimental
repeat_region   complement(4963..5286)
                /rpt_family="L1"
                /evidence=not_experimental
repeat_region   complement(5353..5554)
                /rpt_family="L1"
                /evidence=not_experimental
repeat_region   complement(5542..5625)
                /rpt_family="L1M1"
```

```
repeat_region   /evidence=not_experimental
                complement(5626..5882)
                /rpt_family="AluJb"
repeat_region   /evidence=not_experimental
                complement(5883..6008)
                /rpt_family="L1M1"
repeat_region   /evidence=not_experimental
                6025..6235
                /rpt_family="AluJb"
repeat_region   /evidence=not_experimental
                complement(6236..6447)
                /rpt_family="L1M4"
repeat_region   /evidence=not_experimental
                complement(6486..6554)
                /rpt_family="AluJ/FRAM"
repeat_region   /evidence=not_experimental
                complement(6558..6709)
                /rpt_family="L1"
repeat_region   /evidence=not_experimental
                complement(6778..7721)
                /rpt_family="L1ME2"
repeat_region   /evidence=not_experimental
                9224..9298
                /rpt_family="L1MA6"
repeat_region   /evidence=not_experimental
                9324..9932
                /rpt_family="L1MA7"
repeat_region   /evidence=not_experimental
                complement(8075..8296)
                /rpt_family="L1PA8"
repeat_region   /evidence=not_experimental
                complement(8679..8980)
                /rpt_family="AluSx"
repeat_region   /evidence=not_experimental
                10548..10734
                /rpt_family="MIR"
repeat_region   /evidence=not_experimental
                complement(10013..10508)
                /rpt_family="L1ME2"
repeat_region   /evidence=not_experimental
                11332..11537
                /rpt_family="L1M2"
repeat_region   /evidence=not_experimental
                11540..11728
                /rpt_family="AluSg/x"
repeat_region   /evidence=not_experimental
                complement(10867..11330)
                /rpt_family="L1MA7"
repeat_region   /evidence=not_experimental
                11730..11910
                /rpt_family="L1M2"
repeat_region   /evidence=not_experimental
                complement(11911..12230)
                /rpt_family="L1PA6"
repeat_region   /evidence=not_experimental
                12231..12416
                /rpt_family="L1P"
repeat_region   /evidence=not_experimental
                complement(12640..12746)
                /rpt_family="MIR"
repeat_region   /evidence=not_experimental
                complement(12901..13183)
                /rpt_family="AluSg"
repeat_region   /evidence=not_experimental
                complement(13459..13774)
                /rpt_family="AluY"
repeat_region   /evidence=not_experimental
                complement(13882..14031)
                /rpt_family="MIR"
repeat_region   /evidence=not_experimental
                complement(14171..14238)
                /rpt_family="MIR1E2"
repeat_region   /evidence=not_experimental
                /rpt_family="MIR1E"
                /evidence=not_experimental
```

```
repeat_region   complement(14533..14830)
                /rpt_family="AluSg"
repeat_region   15255..15418
                /evidence=not_experimental
                /rpt_family="MER5B"
repeat_region   complement(15590..15692)
                /evidence=not_experimental
                /rpt_family="AluSg/x"
repeat_region   20210..20503
                /evidence=not_experimental
                /rpt_family="AluJo"
repeat_region   complement(24663..24943)
                /evidence=not_experimental
                /rpt_family="AluJo"
repeat_region   25366..25393
                /evidence=not_experimental
                /rpt_family="(T)n"
repeat_region   complement(25398..25624)
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   25642..25941
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(26664..26814)
                /rpt_family="Aluy"
repeat_region   25945..26235
                /evidence=not_experimental
                /rpt_family="AluYa8"
repeat_region   complement(26283..26574)
                /evidence=not_experimental
                /rpt_family="AluYb8"
repeat_region   29391..29494
                /rpt_family="Aluy"
repeat_region   29495..29772
                /rpt_family="AluSp/q"
repeat_region   complement(29773..30076)
                /rpt_family="AluY"
repeat_region   30113..30372
                /rpt_family="AluY"
```

Query Match

Best Local Similarity 85.9%;
Matches 420; Conservative 0; Mismatches 63; Indels 6; Gaps 2;
Score 356.2; DB 9; Pred. No. 2.1e-24; Length 150724;

```
Qy      18 GAGGACGCTGCCCAGGGCATCGCTAATGAGGACGCCCGCCC              77
           ||||||| ||||||||||||||||||||||||| |||| |
Db   84631 GAGGACGCCCGCCCAGGGCATCGCCAGGGCATCGCCAAGGAGGAC      84690

Qy      78 ACCACCCAGTGCATCGCCAACGAGGAGGAAGCCCCCGCCGCC---GAGGACGCCATC    134
           || | ||||  ||||||||||||||||||||||| ||||||     ||||||||||||
Db   84691 GCCGCCCAGGGCATCGCCAACGAGGAGGTTGCCCAGGGCGCC  GAGGACGCCATC   84750

Qy     135 CAGGGCATCGCCTAACGAGGCCGCGACAAGGCATCGCCACGAGGAGAGCGCCCGCC    194
           |||||||||||| ||  ||||||||||||||||||| | ||
Db   84751 CAGGGCATCGCCAATGGGGTTGCCCCGCGGGG                          84810

Qy     195 ATCGCCAATGAGGACGACGCCCACCCAGGGCATCGCCAACTGTGGACCGTCCACGCGTTCCGCC    254
           ||||||| || ||||||||||| |||||||||||||||||||| || |
Db   84811 ATCGCCAACGAGGACGCCCACCGGCCACCGGCCATCGCCGAGGAGGACGCCATCGCC         84870

Qy     255 AACGGGGACGCCGTCCTCAGCTTCGCCACGGATCGCCACGGGGG              314
           ||| | | ||||| |||||||
Db   84871 AGGGGAGGAGGACGCCCACCGCCCACGGCATCGCCAACGAGGACGCCATCGCC     84930

Qy     315 GACCCCACCAAAGGGCATGGGCAACGAGGTGCCACCGAGGCCAT              374
           ||||||  ||||||||||||||| |||||| || |
Db   84931 GACGCCCATCCAGGGCATCGCCCAAGGAGTACGCCGTTCCACGCGCATCGCCAACGAGGACGCC     84990
```

RESULT 8
AC007324

Db AC007324

```
LOCUS       AC007324      122364 bp    DNA      linear   PRI 18-MAY-2000
DEFINITION  Homo sapiens chromosome 22q11 clone b293, complete sequence.
ACCESSION   AC007324
VERSION     AC007324.55  GI:7923342
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 122364)
AUTHORS     Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A.
TITLE       Homo sapiens Chromosome 22q11 BAC Clone b293 In CES Region
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 122364)
AUTHORS     Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (17-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 122364)
AUTHORS     Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (19-DEC-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 122364)
AUTHORS     Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid,H. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (06-MAY-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 122364)
AUTHORS     Fu,Y., Fang,F., Wang,Q., Pan,H., McDermid.H. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAY-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 18, 2000 this sequence version replaced gi:7712128.
FEATURES             Location/Qualifiers
     source          1..122364
                     /organism="homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22q11"
                     /clone="b293"
BASE COUNT     31795 a   28253 c   28482 g   33834 t
ORIGIN
```

Query Match

Best Local Similarity 85.7%;
Matches 419; Conservative 0; Mismatches 64; Indels 6; Gaps 2;
Score 354.6; DB 9; Pred. No. 3.1e-24; Length 122364;

```
Qy      18 GAGGACGCTGCCCAGGGCATCGCTAATGAGGACGCCCGCCC              77
           ||||||| ||||||||||||||||||||||||| |||| |
Db   82628 GAGGACGCCCGCCCAGGGCATCGCCAGGGCATCGCC--CGAGGACGCCATC    82687

Qy      78 ACCACCCAGTGCATCGCCAACGAGGAGGAAGCCCCCGCCGCC---CGAGGACGCCATC    134
           || | ||||  |||||||||||||||||||||||||||||||     ||||||||||||
Db   82688 GCCGCCCAGGGCATCGCCAACGAGGAGGAGCGCCCGCCGCCCAGGGCATCGCCAACGAGGACGCCCGCCCCAGGGCATCGCCAACGAGGACGCC      82747
```

```
Qy    135 CAGGGCAATCGCCAACGACGAGGAGGTTGCCCAGGGCGATCGCCAATGGGGTCCGCGACAGGGC 194
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  82748 CAGGGCAATCGCCAAGGAGGACGACGCCGTCCCACGGGCATCTCCACGGCTTCGCC 82807

Qy    195 ATCGGCCAATGAGGACGCGCACCCAGGGGCATCGCCAACTGGACGCGCCGTCCACGGCTTCGCC 254
          ||||||| | ||||||||||||||||||||||||||||||||| ||||||| |||||||
Db  82808 ATCGCCAACGAGGACGCCACCCAGGGGCATCGCCAACGGACACTCCGCCCACGGAATCGCC 82867

Qy    255 AACGGGGACGGCCGTCTTCAGCTTCGCCAACGAGGACGCGCGCCCAGGGGCATCGCCAACGGG 314
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  82808 AGCGAGGACGCGCGCCCAGGGGCATCGCCACGGAGGAC 82927

Qy    315 GACGCGCCACCAAGGGCATGGGACGGCTCACCATCCACGGATCGTCAGGGCATCGCCAACGCC 374
          ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  82928 GACGCGCCCAGGGCATCGCCAACGAGGACGCGCCCAGGGGCATCGCCAAGGAGGACGCC 82987

Qy    375 GTCCAGGGCATCGCTAACGAGGTGGCCCCCCAGGGGCATCGCCAACGAGGACGCGCCCAGGGG 434
          | |||||||||| | ||||| |||||||||||||||||| ||||||||||||||||||||||| ||
Db  82988 GCCCAGGGCATCGCCAACGAGGACGCGCCCAGGGGCATCGCCACGGAGGACGCGCCCCAG 83047

Qy    435 GGAATCGCC---GAGGATGTCCCACAGGGGCATGCCCAACAGACGCGCCCAGGGCATC 491
          |  ||||||    |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  83048 GGCATCGCCAAAGAGGACGACGCCGCCCAGGGGCATCGCCAACGAGGACGCGCCCAGGGCATC 83107

Qy    492 GCCAACAAG 500
          |||||||||
Db  83108 GCCAACGAG 83116
```

RESULT 9
AC013360

LOCUS   AC013360   159550 bp   DNA   linear   HTG 04-MAY-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-278E23 map 11, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC013360
VERSION   AC013360.4 GI:7229982
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 159550)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 11, clone RP11-278E23
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 159550)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,P.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 12, 2000 this sequence version replaced gi:6479088.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L2604
            Center clone name: 278_E_23
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 144663 bases at least Q40
            Consensus quality: 153318 bases at least Q30
            Consensus quality: 156484 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 158350; sum-of-contigs
            Quality coverage: 5.4 in Q20 bases; agarose-fp
            Quality coverage: 5.9 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1729: contig of 1729 bp in length
            *   1730   1829: gap of 100 bp
            *   1830   3716: contig of 1887 bp in length
            *   3717   3816: gap of 100 bp
            *   3817   6950: contig of 3134 bp in length
            *   6951   7050: gap of 100 bp
            *   7051   9136: contig of 2086 bp in length
            *   9137   9236: gap of 100 bp
            *   9237   17859: contig of 8623 bp in length
            *   17860   17959: gap of 100 bp
            *   17960   26939: contig of 8980 bp in length
            *   26940   27039: gap of 100 bp
            *   27040   36269: contig of 9230 bp in length
            *   36270   36369: gap of 100 bp
            *   36370   48574: contig of 12205 bp in length
            *   48575   48674: gap of 100 bp
            *   48675   62315: contig of 13641 bp in length
            *   62316   62415: gap of 100 bp
            *   62416   77457: contig of 15042 bp in length
            *   77458   77557: gap of 100 bp
            *   77558   96884: contig of 19327 bp in length
            *   96885   96984: gap of 100 bp
            *   96985   124230: contig of 27246 bp in length
            * 124231   124330: gap of 100 bp
            * 124331   159550: contig of 35220 bp in length.
FEATURES   Location/Qualifiers
     source   1..159550
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-278E23"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..1729
                     /note="assembly_fragment"
     misc_feature   1830..3716
                     /note="assembly_fragment"
     misc_feature   3817..6950
                     /note="assembly_fragment"
     misc_feature   7051..9136
                     /note="assembly_fragment"
     misc_feature   9237..17859
                     /note="assembly_fragment"
     misc_feature   17960..26939
                     /note="assembly_fragment"
     misc_feature   27040..36269
                     /note="assembly_fragment"
     misc_feature   36370..48574
                     /note="assembly_fragment"
     misc_feature   48675..62315

```
RESULT 10

Qy     18  GAGGACGCGTGCCCAGGGCATCGCTAATGAGGACGCCAGGGCATCGCTAATGAGGAC 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36462  GAGGACGCGCGTGCCCAGGGCATCGCCAACGAGGACGCTCCCCACGGAATGCCCACGGAGGAC 36521

Qy     78  ACCACCCAGTGCATCGCCAACGAGGAGAAGCGCCCAGGGCATCGCC---GAGGACGCCATC 134
           | ||||  ||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  36522  GCCGCCCACGGCATCGCCAACGAGGACGCCGTCCAGGGCATCGCCAGGGCATCGCC 36581

Qy    135  CAGGGCATCGCCAACGAGGAGGTTGCCCAGGGCATCCCCAAATGGGGTGGGCCGACAGGGC 194
           |||||||||||||||||||||||||||||||||||||| ||| |||||||||||||||
Db  36582  CAGGGCATCGCCAACGAGGACGCCGCGGCATCGCCAAGGAGGACGCATCGCCAGGGC 36641

Qy    195  ATCGCCAATGAGGAGACGACGCCACCCAGGGCATCGCCAACACTGGGGACGCCGTCCACGGCCTTCGCC 254
           |||||||||||||||||||||||||||||||||||||||| || |||||||||||||
Db  36642  ATCGCCAACGAGGACGCCGCCCAGGGCATCGCCAACGAGGGCATCGCC 36701

Qy    255  AACGGGGACGCGTCCTCAGCTTCGCCAACGAGGACGCCATCGCCCAGGGCATCGCCAACGGG 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36702  AAAGGAGGACGCGCCGCCCACGGGCATCGCCAACGAGGACGCCAACGAGGACGCCAACGAG 36761

Qy    315  GACGCCCACCAAGGCGCATGGGCGTCACCATCCACGGCATCGCTAACGAGGACGCC 374
           ||||| |||| |||||||||||||||||||| ||  |||||||||
Db  36762  GACGCCGCGGAATCGCCACGAGGACGCCCACGGCATCGCCAACGAGGACGCC 36821

Qy    375  GTCCAGGGCATCGCTAAACGAGGTGCCGCCGCCCAGGGCATCGCCAACGAGGACGCCGCCCACGG 434
           ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  36822  GTCCAGGGCATCGCCAACGAGGACGATGCCCACCCAGGGCATCGCCAACGAGGACGCCCAG 36881

Qy    435  GGAATCGC---CGAGGACGTGTCGCACAGGGCATCGCCAACGAGGACGCCAGGGCATC 491
           || ||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db  36882  GGCATCGCCAACGAGGACGCGCCGCCCACGGGCATCGCCAAAGAGGACGCGCCCACGGCATC 36941

Qy    492  GCCAACAAG 500
           ||||| ||
Db  36942  GCCAACGAG 36950
```

ORIGIN
BASE COUNT   42186 a   37364 c   37475 g   41307 t   1218 others

misc_feature   62416..77457
               /note="assembly_fragment"
misc_feature   77558..96884
               /note="assembly_fragment"
               /vector_side:right"
               /clone_end:SP6
misc_feature   96985..124230
               /note="assembly_fragment"
misc_feature   124331..159550
               /note="assembly_fragment"
               /clone_end:T7
               /vector_side:right"

Query Match   70.0%; Score 349.8; DB 2; Length 159550;
Best Local Similarity 85.1%; Pred. No. 7.6e-24;
Matches 416; Conservative 0; Mismatches 67; Indels 6; Gaps 2;